

THIS WEEK



EDITORIALS

SHUTTLE Final flight of *Atlantis* could bring space science down with a bump **p.6**

WORLD VIEW The battle to save Europe's fisheries misses a trick **p.7**

ALARM CALL Nesting birds move home to avoid noisy chipmunks **p.8**



Taken for a ride

Underhand attempts by the US Congress to bypass regulatory decisions made by federal agencies erode science-based policy.

When is science-informed policy not science-informed policy? When US lawmakers saddle must-pass bills with riders that bypass regulatory decisions.

Exhibit A: the removal of the grey wolves in several US states from the Federal List of Endangered and Threatened Wildlife. The Department of the Interior yanked these animals from the list, effective from 5 May, because it was ordered to by Congress, in language tacked onto the appropriations bill for fiscal year 2011 by lawmakers in western states. The wolf was arguably already on its way to delisting — the Department of the Interior wanted to remove it, but was being prevented by the courts. But the wolf's forcible removal from the list at the whim of a handful of politicians sets a dangerous precedent.

Exhibit B: the House version of the still in-process agriculture and food appropriations bill has an amendment that bans the Food and Drug Administration from using its funding to approve genetically modified salmon for human consumption. The Alaska congressman responsible for introducing the amendment was motivated not by aversion to genetically modified organisms, but by straightforward protection of his constituents. He sees the fast-growing salmon produced by AquaBounty of Waltham, Massachusetts, as a potential threat to wild-salmon fisheries in his state.

Are such moves legitimate? The Center for Biological Diversity, a wildlife-advocacy group in Tucson, Arizona, has already sued the government over the delisting of the wolf. The group argues that the move is an unconstitutional breach of US separation of powers, because action in Congress has effectively nullified rulings by the judicial branch.

And Ronald Stotish, chief executive of AquaBounty, said in a statement on 16 June, "Whether or not you support this transgenic salmon, we should all agree these types of shenanigans have no place in a complex scientific debate. These actions threaten the fundamental basis of a science-based regulatory process." The company says that the salmon amendment will be thrown out during the lawmaking process; it may well be right.

Philosophically, there may be nothing inherently wrong with Congress adding such riders to bills. After all, agencies exist to carry out the law as decided by Congress, and if the law changes, so does the business of the agency. There is even a seldom-invoked procedure, laid out in the 1996 Congressional Review Act, that allows Congress to cancel any rule made by any agency by joint resolution, just because it wants to. Ultimately, policy is meant to be the will of the people as expressed through the actions of a representative government.

Just because these riders are above board does not mean they are a good idea. There is no chance that a busy Congress will fiddle with all of the nearly 4,000 rules made each year by US government agencies, but there are always regulations that rise to public notice and offer traction to a lawmaker eager to dash off a rider and add it the nearest massive, essential bill. Congress already has constitutional routes to influence agency decisions, through budget processes and legislation

to frame the function of the agencies. Riders are sneaky short cuts that don't get debated properly and don't deserve to succeed. Whole agencies, from the Food and Drug Administration to the Environmental Protection Agency, were created because the people of the United States, through the medium of their elected representatives, decided that taking a scientific approach to certain decisions was a good idea. Instead of setting soot-emission limits for factories through the legislative branch, for example, the government of the United States, like that of almost every nation on the planet, decided to hand the job to informed experts.

"The people of the United States decided that taking a scientific approach was a good idea."

There is a rising anti-expert zeitgeist among US politicians, which could help to explain the legislative incursions into the business of agency scientists. Perhaps these incursions are also strikes at the political appointees who run the agencies, who are viewed by many as ideologically opposed to a Republican-heavy

Congress because they were appointed by President Barack Obama.

Would objecting to such legal riders be undemocratic — un-American, even? If today's voters prefer to make decisions without scientific input, should their wishes go unopposed?

No. Scientists are voters with a voice too. And most of them probably agree that policy set with scientific input is better than policy set without it.

Agency regulatory processes can be slow. They can be unsubtle and ignore the nuances of circumstances. They are, without exception, highly bureaucratic. But for many questions, from approval of new drugs to the management of fisheries, decisions made by agencies and informed by experts are better for everyone. ■

Activation energy

The US Department of Energy must learn from its shoddy response to misconduct allegations.

The concept of a conflict of interest is crucial to the proper handling of alleged scientific fraud. Investigators appointed to review misconduct allegations should not have close links to either the complainant or the respondent scientists, and the most enlightened institutional policies include clauses to guard against 'real and apparent' conflicts of interest — intelligent phrasing that makes it possible to query the wisdom of an individual's appointment to an inquiry panel without having to question their objectivity.

Most policies guard less well against another kind of conflict of

interest: affiliations to an institution. Managers and administrators have an apparent interest in playing down the extent of any problems at their institutions. A finding of misconduct can result in bad publicity and financial penalties, not to mention the abrupt severance of professional relationships with colleagues who are often well liked and respected.

For taxpayer-funded research, the check on these institutional conflicts of interest should be government oversight. But in many countries, such oversight is patchy or nonexistent. Even in the United States, where a federal policy mandates oversight of institutions by government funding agencies, there are clear and surprising failings — as revealed by a lawsuit that brought to light details of a misconduct investigation overseen by the US Department of Energy (DOE; see page 20). No misconduct was discovered, but errors of judgement were.

The oversight of the investigation had several problems, including that the adjudicating officials did not read the inquiry's report or keep a final copy of it, and that they approved an outcome on the basis of meetings and telephone calls with a lab manager, rather than their own scrutiny of the evidence. The officials who oversaw the investigation were the people who funded the research in the first place — again, there is an apparent conflict of interest.

DOE officials say that their actions are consistent with federal policy, but they seem disconcertingly casual about it. Important decisions were taken informally by a small group of officials and an adviser who apparently shared a common interest: to see the matter quietly resolved. The procedure ought to be more formal, better documented and even adversarial, with the institutional managers required to satisfy officials whom they do not know and who have no stake in the case outcome. This is closer to the more careful oversight of alleged misconduct by both the Office of Research Integrity at the Department of Health and Human Services (DHHS) and the inspector-general of the National Science Foundation.

No wonder policy experts are highly critical of the DOE. One influential US watchdog group, the Project on Government Oversight (POGO) in Washington DC, contacted during *Nature's* reporting of the failures described above, is now preparing a public letter to Steven Chu, the US energy secretary, to argue that the DOE needs a professional office of research integrity. The DOE does have an Office of Inspector

“Officials who adjudicate misconduct cases must take oversight more seriously.”

General, but it typically handles more conventional alleged financial waste and fraud — not scientific misconduct, which requires specialized expertise. The National Institutes of Health (NIH), funded by the DHHS, has a much higher incidence of misconduct allegations in research than does the DOE. POGO suggests — plausibly — that this is the result of a rigorous scientific oversight system that encourages whistleblowers to come forward, rather than of any greater integrity among DOE-funded scientists.

Having an office of research integrity at the DOE could remedy this situation, and might also be politically astute. In a climate in which conservatives are looking for reasons to cut science budgets, and in which energy and climate research is a hot political topic, the DOE needs to be proactive to prevent scandals about alleged scientific misconduct escalating. One needs only to think of the bruising congressional inquiries into NIH-funded science in the 1980s to show the dangers of taking research integrity for granted.

In the meantime, DOE officials who adjudicate misconduct cases must take oversight more seriously. Their goals should be to prevent retaliation against whistleblowing scientists who have raised concerns, to ensure that those who have been unfairly accused are fairly exonerated and to reach the truth about the integrity of the science they fund. A careful look at the evidence is the first step. ■

Up and away

The final mission of the space shuttle heralds difficult days for space science.

The space shuttle is due to lift off on its final mission this week, and many scientists will cheer its departure, both from the planet and from the scene. The shuttle was key to the launch, repair and maintenance of the marvellous Hubble Space Telescope, but its contribution to science otherwise has been mediocre. Scientists often lament that the billions of dollars spent on the programme could have gone on robotic exploration of other planets, space telescopes and similar worthy causes. But they should not be too quick to welcome the demise of the shuttle. Its loss may foreshadow far darker times ahead for all space science.

The shuttle's scientific programme has never lived up to the hype. In its earliest days, programme advocates made outrageous claims that experiments in space might help to tackle HIV and cancer. Recent statements have been more measured and the science more peer-reviewed, but the flight schedule for *Atlantis* this week shows some disappointingly familiar and low-quality space studies: a study of microbial virulence in zero gravity, experiments on weightless mice and an iPhone kitted out with International Space Station apps.

The shuttle was never about science, of course. The Nixon-era programme was designed to give NASA a purpose and, at the time, seemed to be the logical next step: a vehicle that could make fast and cheap excursions to the beginning of space. This has proved more difficult than expected. The programme never flew close to the 50 missions per year originally envisaged, and the cost per flight was always well above the estimates. Nor could the winged orbiter fly solo,

depending instead on bulky boosters and its external fuel tank to give it an extra nudge into orbit — systems that were responsible for the catastrophic loss of *Challenger* in 1986 and *Columbia* in 2003. Each accident revealed flaws in both the design and the operation of the shuttle fleet.

Yet, despite its many shortcomings, it is wrong to condemn the shuttle as a snub-nosed albatross around the neck of research. Like it or not, the space shuttle is probably the most recognized symbol of science and technology for a generation. The shuttle programme was costly, but it kept NASA focused and in the thoughts of both Congress and the public.

The fear must be that without the shuttle programme, NASA and its strong space science will wither. The agency's flagship project, the James Webb Space Telescope, is desperately over budget and likely to be delayed for years. The next Mars rover is also seeing its costs skyrocket, and officials recently stalled plans for a US–European joint mission to the red planet because of mounting budget problems. Efforts to monitor Earth have stumbled with the loss of two high-profile climate satellites in as many years.

What is particularly unfortunate about this current state of affairs is that the possibilities for space science have never been greater. Space-based gravitational detectors could give researchers their earliest look at the Universe, and an array of space telescopes could provide insight into far-flung star systems. Closer to home, landers and rovers could teach us more about the Solar System's history and evolution, and Earth-observing satellites would improve our ability to understand and respond to climate change. Although none of these projects involves an astronaut, all require a strong and vital NASA to succeed.

Eventually, NASA may build another manned space vehicle to replace the shuttle, or it might find a mission that is more in line with the aspirations of the research community. But for now, it is an agency adrift. ■

➤ NATURE.COM
To comment online,
click on Editorials at:
go.nature.com/xhunqy

IFM-GEOMAR



Fishery reform slips through the net

*Upcoming change fails to tackle the pernicious relationship between government advisers and the fishing lobby, says **Rainer Froese**.*

The fishing industry is less important to Europe's economy than its sewing-machine manufacturers. Yet it consistently gets to overrule scientific advice and drive fish stocks to the brink of collapse. Without massive subsidies, European fisheries would be bankrupt: the cost of hunting the few remaining fish would exceed the income from selling the catch.

Every year, fishery scientists are asked by politicians for advice on how hard stocks can be fished without pushing them over the edge, forcing the scientists into a position akin to that of a physician present at a waterboarding session. Adding insult to injury, much of their advice is ignored, and most stocks are outside of safe biological limits.

The sad state of European fish stocks is not a natural or societal failure that good management simply could not overcome. No, it is a desired outcome of Europe's Common Fisheries Policy (CFP), which has allowed the fishing lobby to infiltrate the European Union's political system.

The European public has been led to believe that fisheries management is decided by bureaucrats in Brussels. In fact, member states have the most influence; they make and implement the decisions made in Brussels and control compliance by fishermen. In most member states, this power is with the ministry of agriculture. Yet too many of these ministries have a cosy relationship with the fishing lobby — many of the civil servants firmly believe that it is their job to protect the rights of their national fishing sector, including the rights to obtain subsidies and to overfish. This concentration of explorative, legislative and executive power within a ministry belies what we learned in school about the importance of separated powers in a democratic system.

Perhaps because the public still has romantic notions about fishing, and the media are drawn to fishermen who block ports or dump fish in the streets of Brussels, the European fisheries lobby gets away with destructive tactics. It routinely discredits scientists and their advice, denies the depleted status of the stocks, fights the establishment of protected areas, defends the use of destructive equipment, insists on the right to catch juvenile fish and requests the abandonment of closed spawning seasons and areas. In doing so, it has destroyed the very basis that fishing depends on. European fishers have profit margins of about 3–6%; in New Zealand, which has successfully reformed its fisheries, the margins are closer to 40%.

Fisheries management in Europe culminates in the closed-door meetings of the Council of the European Union. Decisions in the council are typically taken by a two-thirds majority, but need consensus if the European Commission feels that its proposal has been ignored.

That was recently the case in a preparatory meeting on the threatened bluefin tuna. The member states were unhappy with the commission's proposal, which followed the scientific advice, and asked commission officials to leave the room. They then agreed unanimously, with few abstentions, on much higher catches.

Although the ministers change every few years, their advisers remain and oppose any true change. As a result, the CFP has more than 600 regulations, many of which contradict each other. For example, regulated mesh sizes catch smaller fish than the fishers are allowed to land. These fish are then dumped dead at sea. The setting of next year's catches has been described as political horse trading, with unholy alliances supporting each other in an effort to secure the highest possible share for the national fishing sector. Thus, Germany and Poland will support higher French catches in the Atlantic, and France will support

higher catches in the Baltic.

This is the situation that Maria Damanaki faced when she took charge of European fisheries last year. Building on the excellent 2009 green paper on the reform of the CFP, which documents much of what I have described above, Damanaki confronted the council with clear demands for rebuilding European fish stocks until 2015, in accordance with international agreements. The commission's proposal for the reform will be officially published on 13 July.

Considering the mess that we are in, the reform proposes big steps in the right direction. Internationally agreed reference points for sustainable fishing will finally be recognized and the discarding of perfectly good fish for bureaucratic reasons will be phased out. But the proposal falls short of reforms enacted in New Zealand, Australia and

the United States. These countries have precautionary fishing targets and close fisheries when stocks enter the slope to collapse. Europe will have no such precautionary margins and will only gradually reduce fishing pressure when stocks are on the slope to collapse, with no default rule for closing a fishery. Whereas the other countries have phased out or drastically reduced subsidies, the commission proposes only to reshuffle them.

There is no sign that the proposal will tackle the concentration of power with agriculture ministers or the excessive influence of the fishing lobby — for example, by transferring the management of wild fish to environment ministries. The commission's restraint here is understandable: it knows that its plans to save Europe's seas must be approved by the same advisers and lobbyists who have wrecked them. ■

Rainer Froese is a marine ecologist at the Leibniz Institute for Marine Sciences at the Christian-Albrechts University of Kiel, Germany.
e-mail: rfroese@ifm-geomar.de

**THE SAD STATE OF
EUROPEAN
FISH STOCKS
IS NOT A FAILURE
THAT GOOD
MANAGEMENT
SIMPLY COULD NOT
OVERCOME.**

➔ **NATURE.COM**
Discuss this article
online at:
go.nature.com/xvclj8

RESEARCH HIGHLIGHTS

Selections from the
scientific literature

PALAEONTOLOGY

Ancient tooth tells of migration

African ancestors of modern apes first migrated to Eurasia 3 million years earlier than previously thought.

According to fossil records, African 'hominoids' colonized Eurasia around 14 million years ago, giving rise to the evolution of 'hominids' such as orangutans and humans. But a fossilized hominoid tooth found in southern Germany in 1973 suggests that Eurasia's permanent colonization was pre-dated by a less successful wave of hominoid migration.

Madelaine Böhme at the Eberhard Karls University of Tübingen in Germany and her colleagues dated the tooth's age at around 17 million years by analysing the evolutionary state of small mammalian fossils found with it. They then calibrated its age using the magnetic polarity of the surrounding sediment. This corresponds to the direction of Earth's magnetic field, which has reversed occasionally during the planet's history.

J. Hum. Evol. doi:10.1016/j.jhevol.2011.04.012 (2011)

ANIMAL BEHAVIOUR

Birds eavesdrop to skirt chipmunks

Birds that nest low to the ground tune in to the calls of predatory chipmunks in deciding where to set up home.

Ground-nesting birds avoid the territories of egg-eating eastern chipmunks (*Tamias striatus*), but what tips the birds off to predator hotspots has not been

clear. So Quinn Emmering and Kenneth Schmidt of Texas Tech University in Lubbock played chipmunk or frog calls from speakers on plots of land in Millbrook, New York, during two nesting seasons. They found that both the ovenbird (*Seiurus aurocapilla*; pictured right) and the veery (*Catharus fuscescens*; left) built their nests farther from the centre of plots on which chipmunk sounds were broadcast than from those playing frog calls or nothing at all. The effect was most prominent in ovenbirds, which nested an average of 20 metres farther from chipmunk playback stations

than from controls.

This adds to a growing body of evidence that prey animals exploit the communications of their predators — putting this communicative behaviour under selective pressure.

J. Anim. Ecol. doi:10.1111/j.1365-2656.2011.01869.x (2011)

EPIGENETICS

Stressed genes are inherited

Stress in fruitflies can cause changes in gene expression that are passed on to subsequent generations.

A group led by Shunsuke Ishii at the RIKEN Tsukuba Institute in Japan studied the

ecosystem indicators, including the abundance of certain species of zooplankton (pictured).

The team found that biological productivity along coastal California rose when the current shifted north, a trend that is expected to increase as the climate warms. The results suggest that a more northerly current could increase the transport of nutrient-rich subarctic waters into the current that flows south towards California.

Geophys. Res. Lett. doi:10.1029/2011GL047212 (2011)



J. GOMEZ-GUTIERREZ

ECOLOGY

An ecological boost from climate change

Global warming could up the productivity of California's and Oregon's fisheries, thanks to a poleward shift in ocean currents.

William Sydeman at the Farallon Institute for Advanced Ecosystem Research in Petaluma, California, and his group assessed the position of the North Pacific Current — which flows eastwards towards North America, then divides to flow north and south along the coast — from 2002 to 2007. They matched up data from the Argo array of floating ocean sensors with various

protein ATF2 in *Drosophila*. They found that this protein is needed for the formation of heterochromatin — tightly packed DNA-protein complexes that block the expression of certain genes. Stress causes phosphorylation of ATF2, dislodging it from heterochromatin. When *Drosophila* embryos were exposed to heat shock during development, the resulting flies expressed certain genes that are normally silenced by heterochromatin, such as a gene for red eye colour.

The errant heterochromatin packaging and red eye colour were also seen in offspring of males that had been exposed to heat stress during development. *Cell* 145, 1049–1061 (2011)

M. READ/NATUREPL.COM

CANCER

Low carb, low tumour growth

A low-carbohydrate, high-protein diet slowed tumour growth in mice compared with a typical high-carbohydrate Western-style diet.

Gerald Krystal at the BC Cancer Research Centre in Vancouver, Canada, and his colleagues compared the growth of both human and mouse tumours in mice fed diets comprising 8%, 10%, 15% or 55% carbohydrate. All four diets had the same calorie content. At the 10% and 15% levels, mice showed slower tumour growth than animals eating the high-carbohydrate diet, and did not lose weight.

In mice engineered to develop breast cancer, almost half of the animals on the Western diet developed tumours by one year of age, whereas none on the 15%-carb diet did. The authors suggest that for cancers associated with particularly high blood-glucose levels, limiting dietary carbohydrates could aid in treatment.

Cancer Res. 71, 4484–4493 (2011)

ASTRONOMY

Nuclear network detects fireball

An international network of sensors designed to detect illicit nuclear blasts has been used to study the fiery break-up of a large meteoroid over Indonesia in October 2009.

Every decade or so, a meteoroid a few metres in diameter explodes in Earth's atmosphere with a force equivalent to that of a mid-sized nuclear weapon. Elizabeth Silber at the University of Western Ontario in London, Canada, and her group analysed low-frequency sound-wave data from 17 sensors maintained by the Comprehensive Nuclear-Test-Ban Treaty Organization. The team found that the explosion was probably equivalent to between 8 and 67 kilotonnes of

the explosive chemical TNT.

They conclude that the network could be used to study future meteoroid strikes, improving our understanding of near-Earth objects.

Geophys. Res. Lett.

doi:10.1029/2011GL047633 (2011)

CLIMATE CHANGE

Extra warming near Greenland

Rising ocean temperatures are already accelerating the melting of Greenland's coastal glaciers. Projections from 19 climate models now suggest that, with growing greenhouse-gas emissions, the maximum warming of subsurface waters around the island will be almost double the global average increase during the twenty-first century.

Jianjun Yin at the University of Arizona in Tucson and his co-workers modelled the impact of a modest rise in greenhouse-gas emissions on ocean temperatures around Greenland and Antarctica. They predict that waters 200–500 metres below the surface around Greenland will warm by 1.7–2.0°C by the end of this century. Meanwhile, Antarctic Ocean temperatures will increase by roughly 0.5°C.

The authors say that differences in ocean circulation patterns are a key factor in these responses.

Nature Geosci. doi:10.1038/ngeo1189 (2011)

MICROBIOLOGY

A key for bacterial entry

A protein that allows some bacteria to cling to and invade host cells could be a target for combating infection.

Kim Orth and her colleagues at the University of Texas Southwestern Medical Center in Dallas searched for such proteins in the bacterium *Vibrio parahaemolyticus*, which can cause food poisoning. They found MAM7, a protein that is expressed in the bacterial

COMMUNITY CHOICE

The most viewed papers in science

BIOMATERIALS

Barnacles help battle biofilms

HIGHLY READ
on pubs.acs.org
in May

By exploiting barnacles' remarkable clinging ability, researchers have anchored antibacterial 'polymer brushes' to the surface of steel. This could prevent the

formation of dangerous bacterial biofilms on the surfaces of medical devices.

En-Tang Kang at the National University of Singapore and his colleagues coated stainless steel with a thin layer of cement harvested from *Amphibalanus amphitrite* barnacles. They then used this cement to initiate the growth of polymer chains of 2-hydroxyethyl methacrylate, which was coupled with chitosan, an antibacterial molecule. The treated steel lowered the adhesion of the bacterium *Escherichia coli* and cut survival to less than 20%. Normal steel did not affect bacterial viability.

Langmuir 27, 7065–7076 (2011)

NEUROSCIENCE

Protein brings death to myelin

A protein on the surface of certain brain cells can halt production of the myelin sheaths that surround and insulate nerve fibres. The protein, DR6, could be a target of treatments for multiple sclerosis, a disease marked by myelin loss.

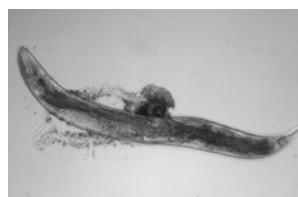
Sha Mi at Biogen Idec in Cambridge, Massachusetts, and her colleagues discovered that overexpression of DR6 causes the myelin-producing brain cells, called oligodendrocytes, to undergo programmed cell death. Antibodies that block DR6 boosted myelin production in cultured oligodendrocytes, and one antibody even reversed myelin damage in a rat model of multiple sclerosis. The post-mortem brains of three patients with multiple sclerosis contained more DR6-expressing cells than brains of people without the condition.

Nature Med. doi:10.1038/nm.2373 (2011)

➔ **NATURE.COM**

For the latest research published by Nature visit:

www.nature.com/latestresearch



MYELIN HAM

FUNDING

California cuts

Universities in California are bracing themselves for more financial pain. The University of California and California State University had each faced a US\$500-million cut in state funds, forcing drastic measures such as the planned closure of a world-renowned library at the Scripps Institution of Oceanography (see *Nature* **471**, 18; 2011). But on 26 June, California lawmakers approved a budget that would cut a further \$150 million from each system. Hikes in tuition fees are likely, and faculty at the University of California say the cuts are compromising its ability to retain and recruit top staff.

European research

The European Commission has proposed that spending for research and innovation should increase to €80.2 billion (US\$116 billion) in the European Union's next seven-year budget, covering 2014–20. That would be a large rise on the current €55 billion (for 2007–13). Research was the only major area to expand in the €1.03-trillion budget proposal announced on 29 June. It will now face scrutiny from Europe's parliament and member states. See page 14 for more.

RESEARCH

Industry ties

Three psychiatrists at Harvard Medical School in Boston, Massachusetts, will face sanctions for not declaring their acceptance of huge payments from drug companies between 2000 and 2007. Joseph Biederman, Thomas Spencer and Timothy Wilens were identified by US senator Charles Grassley (Republican, Iowa) in a 2008

probe into pharma payments. They later disclosed that they had received a combined total of at least US\$4.2 million from drug companies. On 1 July, they told colleagues that Harvard Medical School and Massachusetts General Hospital were banning them from paid industry-sponsored activities for one year, and that they would be monitored for another two years. See go.nature.com/9syihf for more.

Mystery maladies

A US research programme for diagnosing unknown diseases is so overwhelmed with demand that it has stopped accepting new cases, effective from 1 July. In the Undiagnosed Diseases Program, launched in 2008 by the US National Institutes of Health (NIH), expert physicians from across

the NIH try to diagnose mysterious illnesses (see *Nature* **460**, 1071–1075; 2009). The programme's website noted that there is a “substantial backlog” of applications, but officials hope to resume accepting applications in two to three months. Around half of patients seen so far have been shown to have neurodegenerative diseases.

Global plant traits

An international collaboration of plant scientists has published details of its first pass at constructing the world's largest repository of plant traits. The database, named TRY, currently holds almost 3 million entries for 69,000 plant species, and combines contributions from 93 smaller databases (www.try-db.org). It lists 52

groups of traits, including observable biochemical and morphological features such as a leaf's type, size, chemical composition and photosynthetic pathway — all of which, it is hoped, will help to generate reliable predictions of how vegetation boundaries and ecosystem properties shift with climate and land-use changes (J. Kattge *et al.* *Glob. Change Biol.* doi:10.1111/j.1365-2486.2011.02451.x; 2011).

POLICY

Free climate data

The University of East Anglia in Norwich, UK, has been ordered to disclose climate data that it has resisted releasing under freedom-of-information requests. The university had argued that it did not have permission



NASA

Los Alamos lab escapes huge wildfire

New Mexico's largest-ever wildfire stopped short of the Los Alamos National Laboratory — but only just. The Las Conchas fire (pictured from the International Space Station on 27 June) scorched through some 50,000 hectares of land, and encroached on the southwest and

western borders of the nuclear-weapons facility. Fire crews put out one small blaze on the grounds of the lab, which shut down for more than a week. But by 3 July, director Charles McMillan said that the lab had “escaped serious damage”. The facility was set to reopen on 6 July.

J. HOBBS/ALL CANADA PHOTOS/GETTY IMAGES

from all the agencies and organizations that provided the data, which consist of weather readings used to track and calculate climate change around the globe. But the United Kingdom's Information Commissioner's Office has overruled the university, a year and a half after it found that the university's handling of previous similar requests breached UK law. See go.nature.com/cdjflo for more.

French reshuffle

France has a new minister for science and higher education, Laurent Wauquiez, following the promotion of the previous incumbent, Valérie Pécresse, in a cabinet reshuffle. Pécresse's legacy includes reducing state control over French universities. See go.nature.com/8s6mua for more.

Endangered owl

The US Fish and Wildlife Service, which implements the Endangered Species Act, has released a plan to save the northern spotted owl (*Strix occidentalis caurina*, pictured). The species is iconic among conservationists because its 1990 listing as 'endangered' resulted in bans on logging in the forests of the US Pacific Northwest, sparking bitter battles between environmentalists and the logging industry. This week, the agency also confirmed that



long-time staffer Dan Ashe would become its new head, and announced a status review for two bat species: the eastern small-footed bat (*Myotis leibii*) and the northern long-eared bat (*Myotis septentrionalis*). Populations of both species have been hit by the fungal disease white-nose syndrome. See go.nature.com/ipyfmr for more.

Forensics plea

Just as many forensic scientists have warned, the UK government was too hasty when deciding to close down the state-owned Forensic Science Service, a parliamentary committee found in a report published on 1 July. The House of Commons Science and Technology Committee said that the government had overlooked the impact of its December decision on forensic science research and development, and on the nation's criminal

justice system. The report recommends that the closure, planned for March 2012, should be postponed to allow the government time to draw up a forensic-science strategy that includes a national research budget. See go.nature.com/fhjiol for more.

Fracking bans

France has become the first country to ban the practice of hydraulic fracturing, or 'fracking', in which high-pressure fluids are pumped into shale formations to fracture the rock and force out natural gas. The technique has been accused of releasing methane into well water and of polluting groundwater with toxic chemicals. The French parliament vote, on 30 June, came a day after New Jersey's state legislature also voted to ban fracking — the first US state to do so.

Serengeti road

Tanzania has reduced the scope of its plans to build a highway through a northern section of the Serengeti National Park. Researchers had feared that an asphalt highway could destroy the park's ecosystem by interrupting migration routes (see *Nature* 467, 272–273; 2010). They had argued for an alternative route, farther south and outside the park's boundaries. Tanzania's

COMING UP

8 JULY

NASA's space shuttle *Atlantis* launches from the Kennedy Space Center in Florida for its last flight — the final mission of the shuttle fleet (see Editorial, page 6).

go.nature.com/1rszjt

10–12 JULY

In Shenzhen, China, the genomics institute BGI hosts a conference to discuss progress in its project to sequence the genomes of 1,000 plants and animals.

go.nature.com/lhcfkk

tourism minister Ezekiel Maige said that there would still be a new road through the park — in which an iconic wildebeest migration takes place every year — but it would not be paved.

BUSINESS

Avastin aversion

The US Food and Drug Administration (FDA) will probably not reverse its December decision to withdraw approval of the drug Avastin (bevacizumab) for the treatment of advanced breast cancer. Genentech, the drug's maker (owned by Swiss pharmaceutical giant Roche), had challenged that verdict. But on 29 June, an FDA advisory committee voted unanimously that the drug's risks outweighed its benefits. If the FDA commissioner agrees, finalizing the withdrawal, insurers may stop paying for the drug's use against breast cancer. By contrast, the European Medicines Agency declared in December that Avastin's use in advanced breast cancer was "a valuable treatment option".

➔ NATURE.COM

For daily news updates see:

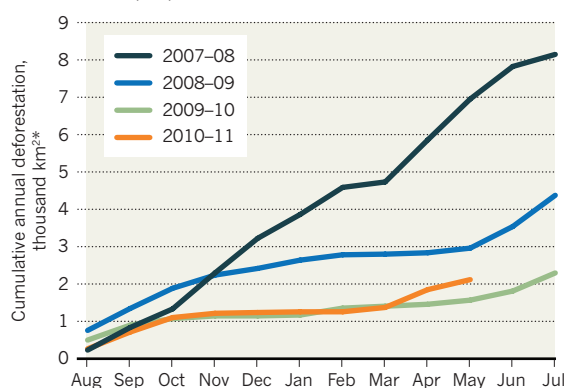
www.nature.com/news

TREND WATCH

Deforestation in the Amazon rainforest has been declining since 2004, but satellite data from Brazil's National Institute for Space Research (INPE) in São José dos Campos, released on 30 June, confirm a trend-busting jump in forest clearing in recent months (see chart). The surge may be related to anticipated changes in legislation that might ease restrictions on clearance. These 'Forest Code' reforms are currently being debated in Congress. See go.nature.com/c51xiC for more.

DEFORESTATION RISES IN BRAZIL

After years of encouraging decline, deforestation in the Brazilian Amazon has leapt up this season.



*As recorded month-to-month by coarse-resolution LANDSAT satellite data

NEWS IN FOCUS

OCEANOGRAPHY Spanish voyage revives a golden age of exploration **p.16**

Q&A Harold Varmus's first year at the National Cancer Institute **p.18**

MALARIA Insecticide resistance threatens to reverse hard-won gains **p.19**

PUBLIC HEALTH Weak science but strong feelings on vitamin D **p.23**



at universities and possibly the research centres too," says neurobiologist Rebecca Matsas at the Hellenic Pasteur Institute, Athens. "But in this economic situation there is no way out, and reforms are urgently needed."

Although the quality of Greece's science is generally seen as poor, the country does boast pockets of excellence. Several internationally competitive research institutes have managed to flourish despite low government investment (see 'Spending gap'). Greek scientists have also been very successful in winning funding from the European Commission's Seventh Framework Programme (FP7) of research, being awarded more euros per researcher than almost any other European Union (EU) country. The planned reforms, spurred by the financial crisis, could help to capitalize on these successes.

The National Council for Research and Technology, a scientific advisory board to the government, is currently designing Greece's first ten-year strategic plan for research. The council was reformed in September 2010 so that most of its members are academics working outside Greece. Using EU structural funds for research, the first major calls for competitive research projects since 2005 are being issued by the government's General Secretariat for Research and Technology. A total of €1.5 million (US\$2.2 million) of these funds must be spent before the end of 2013. "We will be recommending that they are distributed only on merit, through an agency modelled on the US National Science Foundation or the European Research Council," says council leader Stamatios Krimigis, a space scientist at the Johns Hopkins University Applied Physics Laboratory in Laurel, Maryland.

The council's subcommittees, also comprising a majority of scientists working outside Greece, will carry out site visits to evaluate the scientific output of research institutes. These will guide policy on restructuring the research system, and are inevitably causing widespread anxiety. "By next year, there will be fewer institutes," says paediatrician George Chrousos at the University of Athens Medical School, who is also a council member. That will probably happen through mergers, he adds: "The aim is not to fire people but to make them more effective, and the institutes more efficient."

Meanwhile, universities are also set for a shake-up thanks to a new higher-education law, expected to be approved by the end of

Despite protests against austerity measures in Greece, some scientists see an opportunity for change.

POLICY

Greek crisis spurs research reforms

Economic meltdown adds fresh vigour to science shake-up.

BY ALISON ABBOTT

"The threat is very clear," says Greek native Vassilis Pachnis, a developmental neuroscientist at the UK National Institute for Medical Research (NIMR) in London. "There is a danger that this financial crisis could damage the status of centres of research excellence in Greece."

Last week, the Greek government agreed to implement further austerity measures in order to get another international bail-out for its crippled economy. Scientists were already facing tough times when Greece made its first bail-out request in April 2010 (see *Nature* 465, 22; 2010). Their take-home pay, which

was cut by 10% last March, will now shrink by at least another 10% under the new austerity programme. Direct financing of universities and research institutes was cut by about 20% last year, and is set to fall further, although the Greek ministry of education was unable to provide *Nature* with precise figures.

Any damage to science is unlikely to be permanent, Pachnis emphasizes. Like some other Greek scientists, he is hopeful that the country's economic woes may have a positive side. Reforms to the inflexible research and university systems, which do not systematically evaluate research quality or offer competitive grants, are now seen as a necessity in the country's fight for survival. "There is going to be terrible upheaval

► this month. It will scrap a system that has favoured local recruitment, and make it easier for roughly 12,000 Greek scientists working abroad to return. The change may have come just in time: 25% of Greek universities' faculty members are due to retire in the next three years.

Under the law, each university will establish committees (including scientists from abroad) to oversee academic recruitment, which is currently decided by a faculty vote. The law will also modernize university governance, so that rectors will no longer be elected by university staff and students. Instead, a new board of directors will pick their university's rector after an international competition, and will have the power to fire the rector for poor performance.

The next few years will undoubtedly be extremely tough. Although the government says that it is determined to develop Greece's research base, it has not protected research funding in its austerity packages. And it is having trouble meeting its commitments to match research money that scientists have won from FP7. Greece is also negotiating deferrals of its contributions to international organizations such as the European particle-physics facility CERN, near Geneva in Switzerland, and the European Space Agency.

But the most competitive research labs should be sustained by more than €400 million expected to come from EU structural funds by the end of this year. And despite the turmoil, Pachnis is planning to leave the safe haven of his current post as head of the NIMR's neuroscience division to become director of the Institute of Molecular Biology and Biotechnology in Heraklion, Crete. The move offers scientific opportunities that are too good to pass up, he says, and he is confident that Greek science as a whole will survive. "If I can help keep the infrastructure afloat at one centre of excellence through this hard time," he says, "I'll be happy." ■



J. JACQUEMART/EU

EU research commissioner Máire Geoghegan-Quinn (second from left) wants more money for innovation.

POLICY

Europe lines up hefty science-funding hike

Farm subsidies trimmed to enable a 45% rise for research.

BY COLIN MACILWAIN

Like many governing bodies, the European Union (EU) has entered an era of austerity, with one striking exception: scientific research. Under a proposal released on 29 June by the European Commission, the executive body of the EU, spending on research and innovation would rise by about 45%, from €55 billion (US\$80 billion) over the current 2007–13 period to €80 billion in 2014–20. A reduction in farm subsidies would help to pay for science, and spending in other major areas of the €1-trillion budget would remain flat (see 'Innovative thinking').

The proposal is a crucial milestone in a long and convoluted process that will eventually determine the size and format of the successor to the Seventh Framework Programme (FP7) — the EU's flagship research programme — to be called Horizon 2020. The proposal still needs to be agreed by the European Parliament and by member states. But it marks a success for the EU's research commissioner, Máire Geoghegan-Quinn, says Peter Tindemans, head of the science-policy working group at Euroscience, a science-advocacy group headquartered in Strasbourg, France. "She's managed to convince the other commissioners that where there's to be an increase at all, it should be in research."

Other research advocates say that the proposal — although falling short of the major realignment of funding priorities they had been hoping for — was as good as could be expected in the circumstances. "Given the times we're in, we couldn't realistically have

hoped for much more," says Dieter Imboden, president of Eurohorcs, the body representing Europe's national research agencies.

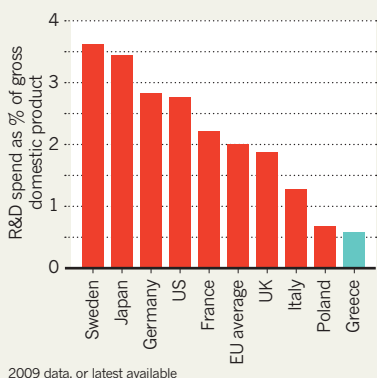
Geoghegan-Quinn told *Nature* that the proposal was "a big vote of confidence in science" but also called on researchers to push to get the proposal implemented — especially in their home countries. "The farmers will be out there lobbying, and scientists and researchers need to do the same," she says.

Commission officials say that the total EU budget for 2014–20 would be worth 5% more, in real terms, than that for 2007–13. The share for research and innovation would grow steadily, and eventually double, from 4.5% in 2007 to 9% by 2020.

The research proposal includes €4.5 billion that would be transferred from farm subsidies

SPENDING GAP

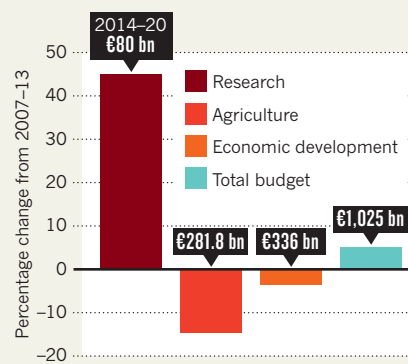
Greece has one of the lowest science investment levels in Europe.



SOURCE: EUROSTAT

INNOVATIVE THINKING

European research funding looks set to grow in 2014–20 under the Horizon 2020 programme, but other budgets will shrink.



SOURCE: EUROPEAN COMMISSION

to pay for agriculture-related research. Although the share of the total budget consumed by the Common Agricultural Policy would fall by some 3%, to about 36%, several observers say that the commission was unable to win support for a more radical transfer of resources from farm subsidies. “The chance to make really big changes to the Common Agricultural Policy has been lost,” says Tindemans.

Geoghegan-Quinn says that there are three main pillars to the proposal: excellence in science, meeting ‘grand challenges’ such as food security, and improving competitiveness.

She says that a much stronger focus will be brought to bear on the grand challenges than before, driven by powerful committees called European Innovation Partnerships. The first such partnership, addressing ‘healthy ageing’, started earlier this year, and is going well, Geoghegan-Quinn says, with its board set to select about ten research priorities, and “push forward with them very quickly”.

Geoghegan-Quinn promises “quick decisions” this year on proposals for the management and structure of the European Research Council, the new and widely lauded component of FP7 that backs basic research, which will be released next week by a commission working group.

The commission’s proposal also says that future cost overruns on ITER, a giant nuclear-fusion experiment under construction in France, will have to be borne by member states, and will not come from the EU budget or fall on other research programmes.

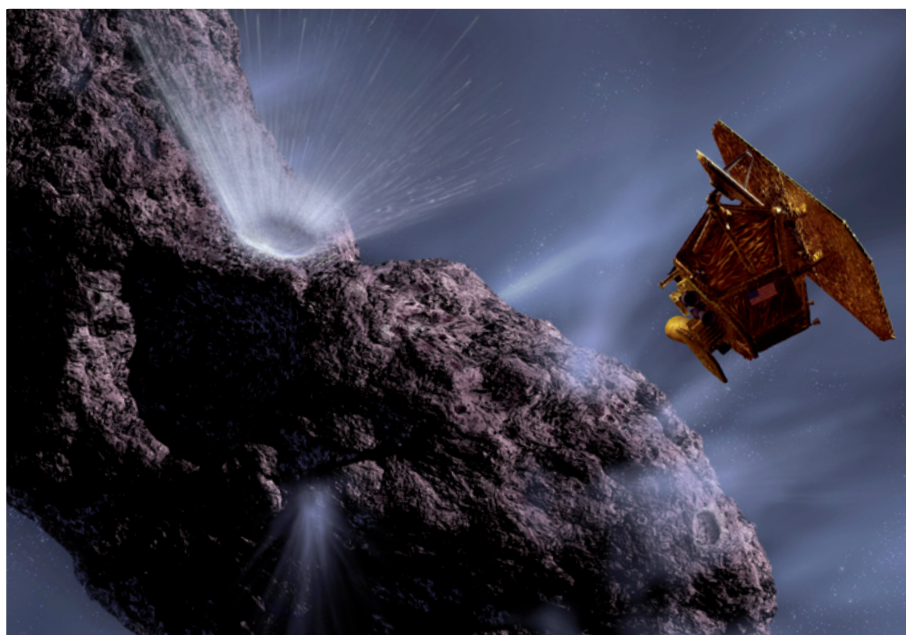
The budget plan will now be scrutinized by the European Parliament and the Council of Ministers, representing the 27 EU member states, before emerging in final form by 2013. The parliament is widely expected to ask for even more funds for research and innovation, whereas many member states will seek cuts to this and other portions of the commission’s proposal.

Jerzy Langer, a physicist at the Polish Academy of Sciences in Warsaw and a close observer of European research policy, says that the outcome so far has been “very good” and indicated “huge determination at the very top” of the commission to support research. He predicted that even such member states as the United Kingdom — which has already said that the overall budget plan is too large — would back the proposed increases for research.

But Luke Georgiou, a policy expert at the University of Manchester, UK, says “it is extremely unlikely that the member states

will agree to anything exceeding this, so we should regard it as a ceiling” on the eventual research budget. ■

➔ **NATURE.COM**
Read more
European stories:
go.nature.com/ptxmzt



NASA/JPL/UMD (ARTIST'S CONCEPT)

Deep Impact had a novice mission leader, but its comet encounter was a smashing success.

PLANETARY SCIENCE

NASA faces dearth of mission leaders

Experience gap looms large in next generation of principal investigators for Discovery programme.

BY ERIC HAND

When NASA invites proposals in 2013 for its next round of low-cost planetary missions, ideas are sure to be plentiful — but not the leaders crucial to the missions’ success. That’s the conclusion of a demographic analysis that shows that the number of highly qualified principal investigators (PIs) willing or able to take the driver’s seat in NASA’s Discovery-class missions is dwindling.

“We have to recognize that this is coming and this is a problem,” says Susan Niebur, who presented her analysis on 21 June at an international conference on low-cost missions at Johns Hopkins University’s Applied Physics Laboratory in Laurel, Maryland. As a cohort of former PIs nears retirement age, Niebur worries that burn out, budget overruns and missed launch windows will be the result if NASA doesn’t find a way to get younger scientists the experience they need to step up into mission-leading roles.

Niebur first encountered the dilemma between 2003 and 2006, when she was the NASA official charged with running the

fiercely competitive Discovery programme — the small, scientist-led planetary probes that are often the most innovative in NASA’s repertoire. At that time, she says, she kept getting proposals “from the same guys”.

Not that they were unqualified. On the contrary, they were precisely the sort of scientists NASA wanted in charge of spacecraft worth hundreds of millions of dollars — people who knew their science but had also dirtied their hands with instrument hardware and experienced the headaches of building a spacecraft. The problem was that there were so few of them — and they were getting older.

Now an independent consultant based in Silver Spring, Maryland, Niebur has been tracking the situation and says that it is getting worse. By 2015, when the winning proposal is chosen, there will be only 14 potential PIs aged 65 and under who have previously been PIs, deputy PIs or project scientists (see ‘Planetary shortfall’). This means that many of the roughly 30 proposals that the Discovery programme attracts at every round will be coming from relative rookies.

Started in 1992, Discovery came to ▶

PLANETARY SHORTFALL

NASA will have few highly experienced scientists available to be principal investigators (PIs) on low-cost planetary missions. But historically, mission leaders have not had high-level previous experience.

Highest level of previous experience		2013 competition (projection)*	Experience levels of PIs, 1994–2011†
High	PI	3	0
	Deputy PI	0	0
	Project scientist	11	0
Medium	Instrument PI	16	7
	Deputy project scientist	11	0
Low	Co-investigator or science investigation lead	50	6
	No mission experience	All others	3

*Scientists expected to be 65 or younger at the time of final selection in 2015.

†Highest-level of previous experience on 16 scientist-led missions, historical or in development.

► embody a new idea: that a space mission could be led by a scientist rather than an engineer. Of the 16 PI-led missions in NASA's history — mainly within Discovery — most have been resounding successes. In some cases, the PIs of those missions stepped into their leadership roles with less experience than NASA anticipated, such as astronomer Michael A'Hearn of the University of Maryland in College Park, who led the Deep Impact mission that pierced a comet's nucleus in 2005.

But that's no reason not to demand more-experienced PIs, says Alan Stern, NASA's science chief from 2007 to 2008. During his stint at the agency, Stern began looking at Discovery, and found that virtually all the projects in development were bursting their budgets.

Some of the PIs had come straight from the blackboard, and some were like “deer in the headlights”, he says. “We’re giving them the keys to the kingdom on projects that are worth hundreds of millions of dollars,” he adds. “That’s an awfully expensive on-the-job training exercise.”

Stern implemented strict rules requiring the heads of all PI-led missions to have had experience as a PI, a deputy PI, an instrument PI, a project scientist or a deputy project scientist on a previous mission. The problem with this was that planetary scientists have few chances to build up such experience. Unlike NASA's other science divisions — astrophysics, Earth science and heliophysics — planetary science does not include balloon and sounding-rocket missions that can give younger researchers a

chance to lead. The Discovery missions, currently capped at US\$425 million and launched every few years, are as cheap and as frequent as they come. Stern's rule was seen as too onerous, and it disappeared when he left the agency.

Rather than ruling out PIs with less experience, Niebur says NASA should enlarge the pipeline of eligible candidates. In particular, she wants the agency to require younger deputy PIs on future proposals. But PIs, trying to keep proposals lean, sometimes exclude deputies.

Michael New, the NASA official now in charge of the Discovery competitions, says that of the 28 proposals being considered in the latest round, which began in 2010, 19 include a deputy PI. However, he says, the deputies are on average the same age as the PIs — suggesting that the demographic problem is not going away.

Nor is it safe to assume that deputies will want to take on the notoriously gruelling task of being a PI. In her analysis, Niebur found that of all the PIs, deputy PIs, project scientists and deputy project scientists on all NASA planetary missions since 1977, none went on to lead another mission as a PI.

That's not too surprising, says Bruce Jakosky at the University of Colorado in Boulder, the PI for the Mars Atmosphere and Volatile Evolution Mission (MAVEN), which was selected in 2008 as the winner of NASA's Mars Scout competition — a programme similar to Discovery. It's just too much work. “I have the heart for MAVEN,” he says, “but not for another one.” ■

SOURCE: S. NIEBUR

OCEANOGRAPHY

Spain's ship comes in

Globe-trotting expedition hopes to prove the value of old-fashioned scientific seafaring.

BY LUCAS LAURSEN

In the age of networked buoys and remote-sensing satellites, a global oceanographic cruise might sound like a relic from the golden era of exploration.

But the seven-month trek of Spain's BIO *Hespérides*, which concludes next week when it docks in Cartagena, aims to deliver a global, comprehensive portrait of the ocean and how it is changing that the project's backers say could not be assembled in any other way.

The Malaspina expedition, organized by Spain's National Research Council (CSIC), set out on 15 December last year. Named after Alessandro Malaspina — who led a five-year survey of the Spanish empire's natural history, economy and geography in the late eighteenth century — the €17-million (US\$25-million) effort unites contributions from the Spanish

Ministry of Science and Innovation, the Spanish Navy and the BBVA Foundation, the charitable arm of the BBVA banking consortium.

Most oceanographic voyages are shorter than the Malaspina expedition (see ‘Voyage of discovery’). By pouring resources into a single voyage, Malaspina's planners hoped to gather global data sets in a wide variety of research areas, from the distribution of persistent organic pollutants to the discovery and genetic characterization of deep-sea life. They also wanted to follow the progress of carbon from the atmosphere as it sinks through the sea, gets captured in the planktonic food chain and eventually comes to rest on the ocean floor. And the scale of the effort has helped to attract international collaborators. Carlos Duarte, a marine biologist at the Mediterranean Institute for Advanced Studies in Esporles who led the mission, says that the approach

was economical because loading instruments for one project costs less than changing instrumentation every month, as the research vessel would normally do.

“There are several other ocean exploration cruises going on, but Malaspina is a much more substantial scientific effort,” says marine biologist Larry Madin, director of research at the Woods Hole Oceanographic Institution in Massachusetts, who is not involved in the expedition. For instance, the Atlantic Meridional Transect is a 16-year-old British project that is making comparable multidisciplinary observations, but is restricted to the Atlantic.

The few cruises with a global reach are mainly sailing boats, such as Craig Venter's Sorcerer II Expedition or the French Tara Oceans expedition,

➔ **NATURE.COM**
For more from the
Malaspina cruise,
see:
go.nature.com/dwlp2w



The *Hespérides* (above, docked in Cape Town) has gathered samples and data during a seven-month global cruise.

which carry fewer sampling instruments and limited on-board research facilities.

By contrast, the 82.5-metre-long *Hespérides* lumbered across the sea bearing laboratories staffed around the clock by two dozen graduate students and veteran researchers from many disciplines. Investigators analysed each day's instrument data and plankton catch on-board, allowing them to devise new observations on the fly. After Japan's Fukushima nuclear accident in March, for example, the team began tracking radiation levels in its seawater samples. The researchers also tracked temperature, salinity and levels of dimethyl sulphide, a plankton waste product that may influence cloud formation above the ocean. And they froze viral and bacterial samples taken from a depth of 4 kilometres, planning to sequence the microbes' genomes.

Duarte says that the modern-day Malaspina expedition achieved a major objective before it even left port: it persuaded hundreds of oceanographers to agree on common research objectives and methods. This approach should allow scientists to compare cruise data wherever they were gathered, says marine biogeochemist Eric Achterberg of the National Oceanography Centre in Southampton, UK, who is not part of the expedition. He and other researchers say that conflicting observational protocols often prevent researchers from comparing data sets taken by different teams in different oceans. Future expeditions could use the Malaspina

protocols as a model. "We're making the same observations around the world," says María de Oca Echarte, a researcher at the Mediterranean Institute for Advanced Studies who took part in Malaspina.

The expedition should also fill gaps in global oceanographic data. Some regions visited by the *Hespérides*, such as the southern Indian Ocean and parts of the southern Pacific, see

during the cruise. Investigators found that some plankton was sinking to the depths much faster than had been previously observed. If researchers confirm the measurements, and if such sinking is widespread, the discovery could affect estimates of how fast the ocean can sequester carbon.

But several scientists contacted by *Nature* say they are waiting to see more of Malaspina's data before they decide whether it has met its goals. The expedition won funding that could have gone to projects in other disciplines, and occupied all of the *Hespérides*' research time this year, so Spanish researchers will scrutinize its results and debate whether to conduct future oceanographic research projects along the same lines.

Physical oceanographer Álvaro Viúdez of the CSIC's Institute of Marine Sciences in Barcelona says that the overarching goal of evaluating "global change" in the ocean is "an exaggerated objective, outside the capacity of an expedition of this type". He adds: "The most relevant

data for evaluating global change in the ocean are those collected by the thousands of Argo buoys, which have been collecting data for years, together with satellites."

Viral molecular ecologist Matthew Sullivan, a collaborator on the Malaspina expedition at the University of Arizona in Tucson, says that time will prove the expedition's worth: "Buy-in will happen a year or so from now when papers start to come out. People will see the value." ■

VOYAGE OF DISCOVERY

The circumnavigation of the BIO *Hespérides* took in less-studied areas of the Pacific and Indian oceans.



fewer oceanographic expeditions than do waters closer to the powerhouse nations of oceanographic research such as Germany, Japan, the United States and Britain. Madin says that the cruise's data could keep researchers busy for many years, noting that results from HMS *Challenger*, a pioneering research cruise in the late nineteenth century, are still used by scientists today.

A few preliminary results emerged



Q&A Harold Varmus NIH cancer chief wants more with less

Harold Varmus, the high-profile director of the US National Institutes of Health (NIH) from 1993 to 1999, returned to the biomedical agency last July as director of the National Cancer Institute (NCI). In 1989, Varmus shared the Nobel Prize in Physiology or Medicine for his studies on the genetic basis of cancer. More recently, he headed up the Memorial Sloan-Kettering Cancer Center in New York. A year after taking the reins at the NCI, Varmus spoke to Nature about his latest role and about the disease that has defined his career.

You had already headed the entire NIH. Why did you take the job as NCI chief?

When I was the NIH director, I often expressed envy of institute directors: they had the money and ran the scientific programmes. I was right — this job is more interesting.

What have been particular satisfactions or successes during your first year at the NCI?

Refurbishing the leadership team at the NCI, pursuing some important initiatives in cancer genomics and global health, identifying 'provocative questions' — important but neglected questions about many aspects of cancer — and working with some extraordinary colleagues. It is also a pleasure to be in place when important projects come to fruition: the National Lung Screening Trial, the pilot phase of the Cancer Genome Atlas, therapeutic trials for metastatic melanoma, and many smaller-scale studies of cancer biology. And I am happy to return to the best urban bike commute in the US: 12 miles through Rock Creek Park.

The NIH is facing its toughest budget in decades. How has that affected your first year at the NCI?

Some effects of this year's budget are self-evident. An actual decline in real dollars — unprecedented in my time in government — has slowed the progress that is so important to the public, caused a lot of distress in our scientific community, and required extra attention from many NCI staff, including me. I have tried to make fair decisions without losing sight of the fact that we do have a US\$5-billion budget and should be able to do many things, including some new things, with those funds to take advantage of unparalleled scientific opportunities.

The public has high hopes that investment in the NCI will lead to cures for cancer. Can you deliver?

Hope is essential for public support, but it has to be framed around the reality of cancer, not founded on simplistic concepts of 'the war on

cancer' or 'a cure for cancer'. Cancer is a collection of many diseases with common principles, and each disease will have to be understood and more effectively controlled on its own terms.

What should basic scientists know about your intentions and priorities for the institute?

All basic scientists who look to the NCI for funding should know that I will tolerate no retreat on the study of model systems and the pursuit of fundamental biological principles. We have just begun to catalogue genetic and epigenetic changes in cancer cells and to probe the significance of those changes. To make optimal use of that information, we need a more profound understanding of cell and developmental biology.

At a time of historically low grant-application success rates, what are you doing to protect investigator-initiated research?

We are attempting to fund about 1,100 new grants this year by making small reductions in most components of our budget; because the number of applications remains high, the success rates will be relatively low. Under these circumstances, we do not have a traditional, sharp payroll. The NCI scientific programme leaders meet regularly to ensure that we are not ignoring highly original proposals and that we are not creating an unbalanced grant portfolio.

What would you say to those who fear that the NIH is moving away from investigator-initiated research and towards large-scale projects?

I don't see much evidence for the claim that the NIH is 'moving away' from investigator-initiated research. It is the largest component of the NCI's portfolio and the one that I am working hardest to protect. But there are programmes that need to be undertaken as community efforts because they are expensive and difficult, yet promise advances: the renovated clinical-trials system, genomic studies and certain translational activities. This does not mean taking away responsibilities from our investigators; it means working closely with the relevant investigators to ensure that we develop large-scale projects in a sensible way.

What is your sense of the public's understanding of cancer science and medicine?

I think it is becoming more sophisticated, especially as new genetic findings and the first targeted therapies are more widely recognized. We are also helped by Siddhartha Mukherjee's prize-winning book, *The Emperor of All Maladies: A Biography of Cancer*. The book explains cancer research sensibly and forcefully, and it portrays cancer in ways that show both how far

we have come in treating some cancers and how much more can be done.

➔ NATURE.COM

For a longer version of this interview, visit:
go.nature.com/akwzou

**INTERVIEW BY
MEREDITH WADMAN**

DISEASE CONTROL

Mosquitoes score in chemical war

Growing resistance is threatening global malaria-control efforts.

BY DECLAN BUTLER

Key weapons in the fight against malaria, pyrethroid insecticides, are losing their edge. Over the past decade, billions of dollars have been spent on distributing long-lasting pyrethroid-treated bed nets and on indoor spraying. Focused in Africa, where most malaria deaths occur, these efforts have greatly reduced the disease's toll. But they have also created intense selection pressure for mosquitoes to develop resistance.

"Data are coming in thick and fast indicating increasing levels of resistance, and also of resistance in new places," says Jo Lines, an entomological epidemiologist and head of vector control at the Global Malaria Programme of the World Health Organization (WHO) in Geneva, Switzerland. The WHO now intends to launch a global strategy to tackle the problem by the end of the year.

Pyrethroids are the mainstay of malaria control because they are safe, cheap, effective and long-lasting. Alternatives such as organophosphates and carbamates are available for indoor spraying, although they cost more and are less effective. But pyrethroids are the only insecticides approved by the WHO for use in bed nets. "We have lots of our eggs in the pyrethroid basket," says Robert Newman, director of the Global Malaria Programme.

The international community has been slow to respond to the threat despite warnings, says Janet Hemingway, director of the Liverpool School of Tropical Medicine, UK, and chief executive of the non-profit Innovative Vector Control Consortium, a public-private venture set up in 2005 to develop new insecticides and monitoring tools. "A number of us had been banging the drums, saying: 'As soon as you scale up you are going to get resistance.'" But

Lines says that the malaria-control community felt too many lives were at stake to let the threat of resistance stand in the way of massively scaling up the bed-net and spraying campaigns.

Teasing out the impact of resistance on the success of malaria-control interventions is difficult because so many other factors influence their outcome. More systematic and more sophisticated monitoring of resistance is also vital, says Lines. The best surveillance data

increased levels of mosquito enzymes that can destroy pyrethroids before they reach their target, require more complex tests to detect (H. Ranson *et al. Trends Parasitol.* 27, 91–98; 2011).

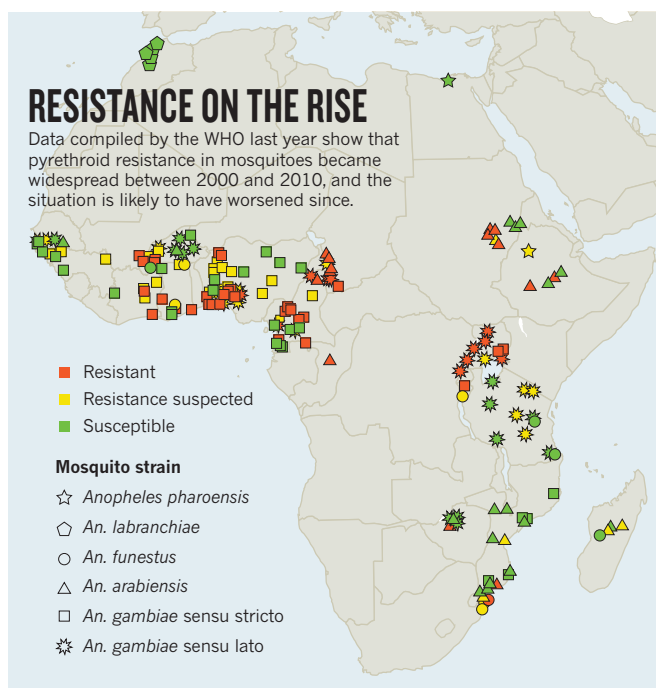
But uncertainties about the extent of resistance or its impact are "no excuse for inaction", says Newman, arguing that the proposed WHO strategy needs to be urgently implemented, and also rolled out preemptively in places where resistance has yet to be detected.

The WHO's plan will recommend, for example, that control programmes rotate insecticides sprayed indoors, using pyrethroids one year and a different class the next. This would be more costly and less effective than relying only on pyrethroids, however, so control programmes may be reluctant to adopt this measure.

Lines says that new combinations of insecticides also need to be developed, so that mosquitoes resistant to one would be killed by the other. In areas where pyrethroid bed nets are used, a different class of insecticides should be used for wall spraying, he adds.

Ultimately, entirely new classes of insecticides — particularly those that can be applied to bed nets — are needed to alleviate the dependence of malaria-control efforts on pyrethroids. For indoor spraying, some longer-lasting and more cost-effective non-pyrethroid insecticides should be available by next

year, Hemingway says, although developing wholly new classes will take five to seven years. Repurposed agricultural insecticides might also act as a stopgap were resistance to pyrethroids to develop rapidly. Research targeting mosquito control is "grossly underfunded" compared with that on malaria drugs and vaccines, she adds, which is why control efforts have had so few options to call on. ■



(see 'Resistance on the rise'), although useful, do not give a complete picture of where resistance is emerging and how prevalent it is, he says. Malaria-control programmes often lack insect-resistance monitoring, and detection of all forms of resistance is not easy. Quick, cheap tests can pick out gene mutations that help the mosquitoes' nerve cells withstand pyrethroid attack. But other forms of resistance, which depend on

MORE ONLINE

TOP STORY

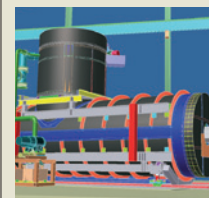


Survey reveals treasure trove of rare-earth elements on sea floor
go.nature.com/tm23gy

OTHER NEWS

- Technique provides clues to the shading of fossil animals go.nature.com/n4dhmh
- Oil-spill aftermath hampers rig research go.nature.com/tyzhoy
- Better biosurveillance could halt disease spread go.nature.com/5sqj33

ON THE BLOG



The top five efforts in fundamental neutron physics
go.nature.com/zouq58

INVESTIGATION

CLOSED

A legal case about public access to documents is raising questions about the US Department of Energy's scrutiny of alleged scientific misconduct.

BY EUGENIE SAMUEL REICH

On 6 April, a federal district judge in Boston, Massachusetts, dismissed a lawsuit that I had filed in 2009 under the US Freedom of Information Act. He concluded that the US government does not have to release a report on an investigation into a case of alleged scientific misconduct at a national laboratory. The ruling was disappointing but liberating: I finally had occasion to write about a case that has shown how the US Department of Energy (DOE) takes a strikingly hands-off approach to the oversight of such investigations.

The lawsuit concerns a report on an investigation into vehemently denied allegations of data fabrication by scientists who receive millions of dollars per year from the DOE's Office of Science. I have asked the judge to reconsider his decision, so the case remains pending. But the larger issue of lax oversight is highlighted in documents filed in court by the government, including nine sworn declarations from DOE and lab officials. These documents show that the DOE officials overseeing the investigation merely skim-read an incomplete version of the resulting report, did not read the complete final report, did not keep a copy of that final report anywhere in their files, and voluntarily gave away a version that was sent to them. I have also learned, through interviews, that scientists who tried to contact the DOE to express concerns about the investigation did not get a hearing.

The court filings suggest that agency officials did not think that they were acting improperly; indeed, they justify their actions by citing the US federal policy on research misconduct, which since 2000 has governed how

allegations should be handled. But the DOE's approach contrasts with much stricter interpretations of the same policy at some other US government science-funding agencies, which typically handle such allegations in an independent office staffed with professional investigators. The DOE is unusual in allowing the allegations to be handled by the officials who awarded the grants in the first place, who will potentially be reluctant to see a problem in that research. Independent observers worry that the DOE's loose oversight could undermine the credibility of research funded through its Office of Science, which supports 10 national laboratories and around 27,000 scientists.

CORRECTIONS AND REFUTATIONS

The allegations that were under investigation first emerged in 2006, when an anonymous peer reviewer accused a research group at Oak Ridge National Laboratory (ORNL) in Tennessee of fabricating data in two manuscripts: a then-current submission to *Nature Physics*, which has not been published, and a paper that had been published in 1993 (*Nature* **366**, 143–146; 1993). The ORNL group is headed by Stephen Pennycook, an electron microscopist, and the earlier work had been a landmark paper in atomic-scale imaging. Pennycook, a pioneer of such techniques, has

been pushing the limits of spatial resolution in electron microscopy to solve problems in a variety of research areas, including materials sciences, nanotechnology and condensed-matter physics.

NATURE.COM
To read more
about misconduct
investigation, visit:
go.nature.com/bxcnbn

Pennycook told me in an e-mail that he preferred not to comment for this article, “as I think there is nothing to discuss”. He does address the allegations on the group’s website (go.nature.com/gtyqpm), where he denies that his group fabricated or falsified data — although he admits that the researchers made errors of judgement in their data presentation. A panel of investigators appointed by ORNL managers upheld the group’s conduct. In a summary statement posted on the ORNL website in 2008 (go.nature.com/ilywxc), the three scientists on the panel said that there was evidence that Pennycook’s team had made careless factual errors and errors of judgement in data presentation, but there was no evidence for misconduct or fraud. In an interview with me, David Williams, one of the investigators and now dean of engineering at Ohio State University in Columbus, went further, slamming whoever made the anonymous allegations. “To go around claiming fraud is a witch hunt of zealots,” he said. “It doesn’t serve the cause of science.”

Following the investigation, the Pennycook group published a correction to its 1993 paper. Nonetheless, some scientists outside ORNL were uncomfortable with the investigation, and particularly with the managers’ failure to release the resulting report. “It’s taken place in the dark,” says John Silcox, an electron microscopist at Cornell University in Ithaca, New York, who was a reviewer of the 1993 paper. It was also unclear how officials at the DOE, which funds Pennycook’s group to the tune of around US\$2 million a year, had overseen the investigation.

This lack of transparency at both ORNL and the DOE was a leading motivation for my freedom-of-information request to see the report. Although the report has not been released, declarations filed in court in response to my lawsuit have shed light on the oversight of the investigation.

In 2006, the declarations say, James Roberto, then ORNL deputy director for science and technology, sent a report on the lab’s investigation to Patricia Dehmer, then the associate director of basic energy sciences in the DOE’s Office of Science, and now the office’s deputy director. In her own declaration, Dehmer says that she understood that this report was only a draft, because it said as much on the first page, and because it didn’t contain referenced appendices — although she does not

however, Dehmer didn’t read the report at all. Her declaration to the court says, “a copy of briefing materials in a binder was sent to me. I do not recall the contents of this binder, nor do I recall reading or studying these materials ... due to the other pressing responsibilities during this time frame.” Later, she and several members of DOE staff went to a meeting with Roberto, where they questioned him about the case. Dehmer approved the investigation, she says, on the basis of the “thorough” responses that she received from Roberto.

In declarations to the court, five DOE employees who attended the meeting say that they did not read the final

“TO GO AROUND CLAIMING FRAUD IS A WITCH HUNT OF ZEALOTS. IT DOESN’T SERVE THE CAUSE OF SCIENCE”

report. At the end of the meeting, Dehmer gave the document to Roberto to take back to ORNL; nothing was left in the government’s possession except the incomplete e-mailed copy that Dehmer had skim-read.

This procedure raises substantial concerns among external experts. Christine Boesz, a former inspector-general overseeing research integrity at the National Science Foundation (NSF), says that it is highly problematic for the government not to keep possession of a final report about a case of alleged misconduct. She points out that government agencies need to keep records, in case of a change of personnel. A failure to fully document cases could make it impossible for the agency to police its research effectively, because it wouldn’t be aware of patterns or alleged patterns involving contractors. “Even a pattern of allegations can tell you a lot,” says Boesz. “The written document is an integral part of the process. If this is the way they typically handle it, then it raises big questions in my mind. Is this looking after the government interest?”

Sybil Francis helped to draft the federal policy on research misconduct while working at the White House Office of Science and Technology Policy in 2000. She says that, according to the spirit, intent and letter of that policy, the government should have access to the documents it needs to conduct oversight. “I don’t see how a federal agency can carry out its oversight responsibilities in a case of alleged or proven research misconduct without the proper documentation at its disposal,” she says. Francis adds that at the very least, she would expect an adjudicating official to get a member of her staff to read the investigation report carefully, and not simply rely on oral presentations.

C. K. Gunsalus, a law professor who studies research misconduct at the University of Illinois at Urbana-Champaign, is perplexed by DOE officials’ failure to read the report. “I think it’s breathtaking, I’m truly astonished,” she says. She contrasts the DOE’s practices with those of two other government agencies: the NSF and the Office of Research Integrity at the Department of Health and Human Services, which funds research at the National Institutes of Health. “The submission of an investigation report to the Office of Research Integrity or the inspector-general of the NSF means without question that the report will be examined with care,” she says. She notes that it is not uncommon for an investigation committee’s first draft

“I DO NOT RECALL THE CONTENTS OF THIS BINDER, NOR DO I RECALL READING OR STUDYING THESE MATERIALS”

specify what she would have expected such appendices to consist of. Dehmer says that she looked through the document only briefly, to check that it reflected what Roberto had told her about the case during conversations and telephone discussions.

In an e-mail, Dehmer told me that by November 2006, ORNL and the Office of Science considered that the investigation was complete. But some time after that, ORNL reopened the case, because managers there had received extra material, the nature of which they do not describe in the court documents. As a result, ORNL produced an extended final report; the lab sent this revised document to the DOE by overnight post. This time,

to flinch from difficult conclusions, to make statements without support or to overlook something. But at these other federal agencies, oversight is iterative, with officials encouraging a thorough investigation by examining the evidence for themselves and asking investigators at the institution for further documentation. If they don't read the report, they can't do that.

Peter Stockton, a senior investigator at the watchdog group Project on Government Oversight in Washington DC, has studied management of weapons labs by the DOE. He says that a general problem at the labs and at DOE headquarters is that officials do not want to probe difficult situations that might result in political fallout and budget cuts. "The most important thing is to limit damage to the organization," he says. He comments that the Office of Science has a large enough budget, at \$4.9 billion a year, for the DOE to consider launching an office of research integrity to oversee it.

Dehmer would not comment on these criticisms. But in her declaration filed in court, she says that federal policy allows government agencies to delegate the job of investigating to research institutions, because to do otherwise "would have involved a substantial new federal bureaucracy, which is not thought desirable" — a quote from the federal misconduct policy. The current director of the Office of Science, William Brinkman, also did not respond to my request for comment, and a DOE spokeswoman notes that "the department does not comment on specific questions relating to litigation matters".

Raymond Orbach, director of the DOE Office of Science at the time of the investigation, says that he does not recall any details of the case. In general, he says, he reads investigation reports when acting in oversight of a case. Orbach notes that ethics are usually handled by the general counsel's office at the DOE, and are a very serious issue for any government agency. But he disputes the value of an office of research integrity at the DOE, saying that he is concerned about the possibility of oversight boards probing into mistakes that scientists have made. "The community has a well-worn method for policing itself," he says: checking whether results are reproducible.

EXTERNAL INTEREST

While the DOE was preparing to close the case, two electron microscopists at Cornell, David Muller and Silcox, were undertaking an independent analysis of the Pennycook group's 2006 correction. In 2006, the two researchers submitted a technical comment for publication in *Nature*, claiming to show that Pennycook's correction is inaccurate, and that the central claim of the 1993 work — that the team had successfully used an electron microscope to image the interface between two materials with atomic resolution — is not supported. Neither Muller nor Silcox made the original allegations against Pennycook's group, although Muller says that he gave advice to the scientist who did. *Nature* accepted the technical comment for publication, pending a reply by Pennycook. A draft of that reply says that the technical comment contains "a mix of criticism with little scientific basis, false statements, irrelevancies, and suggestions of scientific misconduct", and that the group has been cleared. Although Muller's comment (J. Silcox and D. A. Muller Preprint at <http://arxiv.org/abs/1106.4534>; 2011) was accepted, it was never published in *Nature*. Karl Ziemelis, *Nature*'s chief physical sciences editor, told me that he cannot comment on unpublished submissions, but notes that, in general,

technical comments are a forum for scientific discussion, not for airing allegations of misconduct.

The DOE was aware of Muller and Silcox's analysis; in 2007, Dehmer referred to it in an e-mail to me. But the agency has never asked to see it, and when Muller and Silcox tried tentatively to raise their concerns with the relevant officials, they had a frosty reception. Muller says that his initial approach was to contact ORNL to express his concerns and offer full details to any independent committee, but none was appointed. Barbara Penland, a spokeswoman for ORNL, says that any extra concerns that the lab received about the case were referred to an internal inquiry that did not find any issues that warranted

**"IF THIS IS THE WAY THEY
TYPICALLY HANDLE IT, THEN
IT RAISES BIG QUESTIONS IN
MY MIND"**

further investigation, and that the entire ORNL response was assessed by the science and technology committee of the board of governors of UT-Battelle, the contractor that runs the lab. "It concluded that the investigation was handled properly, free of conflict of interest and consistent with applicable federal guidelines," she says.

Muller and Silcox did consider contacting the DOE directly. Cornell's vice-provost for research, physicist Robert Richardson, called Orbach to try to open a line of communication, but was unsuccessful. Orbach, who is now director of the Energy Institute at the University of Texas, Austin, says that he does not recall specifics of any phone calls. Silcox says that he tried to contact Harriet Kung, current associate director of basic energy sciences at the Office of Science, who was involved in approving the investigation according to the government filings, but that she would not discuss the case with him. Kung did not respond to a request for comment.

For Muller, the lesson of this episode is clear. "The DOE needs an office of research integrity," he says. Not only would that allow for independent oversight, but it would also protect complainants who risk losing their own funding if their concerns are not welcomed by grant officers. Since he submitted his comment to *Nature*, Muller has himself received DOE funding, and he admits that the fear of losing that money makes it harder to bring his concerns forward.

As for my lawsuit, the DOE's handling of the investigation report seems to have protected the document from disclosure, at least for now. The judge found that the report is not a government record releasable under the Freedom Of Information Act — primarily because government officials have never read it. I have brought a motion to reconsider, so the case remains open. But an unexpected question has now been raised: should taxpayers be allowed to read about how alleged misconduct at the US national labs has been investigated, even if government officials don't? ■ **SEE EDITORIAL P.5**

Eugenie Samuel Reich is a contributing correspondent for *Nature*. In *Reich v. the US Department of Energy 1:09-cv-10883-NMG* in the US District Court in Boston, Massachusetts, she is represented by David B. Smallman and Michael A. Pezza Jr.



The vitamin D-lemma

A vociferous debate about vitamin-D supplementation reveals the difficulty of distilling strong advice from weak evidence.

BY AMY MAXMEN

With his skull-and-crossbones bow tie tied tight, Clifford Rosen strides to the podium at the Metropolitan Bone Club, a meeting of researchers and clinicians in New York City concerned with all things skeletal. He begins by bracing himself: "If you want to ask a question or just yell at me, go ahead," he says. "I'm used to a lot of antagonism, anger, and frustration."

Rosen is director of clinical and translational research at Maine Medical Center Research Institute in Scarborough and is a respected member of the bone-research community. But his role last year on an expert panel to determine how much calcium and vitamin D people need put him at odds with many of his colleagues. In the past few years, vitamin D has earned a reputation in Western countries for preventing or fighting prostate cancer, cardiovascular disease, multiple sclerosis and about 30 other maladies, leading to advice that most people should be supplementing what the body produces naturally when exposed to sunlight. But in November, the panel, put together by the Institute of Medicine (IOM) — a nonprofit group affiliated with the US National Academy of Sciences — issued a report¹ that

challenged that view. Blood levels of vitamin D need not be as high as many physicians and testing companies had been advocating, it said, and high doses of the vitamin could actually cause harm. Since the report was released, Rosen says he's received about 150 e-mails critical of the panel's decisions. About one-third were downright hateful. "A rehabilitation doctor in Texas threatened to bring me to the board of malpractice to have my licence revoked. People tell me I don't know what I'm doing," he says. "It has become personal."

Much is at stake. By 2009, the amount spent on vitamin-D supplements in the United States had risen tenfold in ten years (see 'Raising the stakes'). Medical practitioners and public-health officials worldwide look to the IOM for guidance on how to interpret the conflicting claims about vitamin D. Yet several vitamin-D proponents say that the IOM's methods, which involved a systematic review of the literature, were flawed. They have accused the panel of misinterpreting data and over-emphasizing the danger of heavy supplementation. Just last month, the Endocrine Society, a professional association of 14,000 researchers and clinicians based in Chevy Chase, Maryland, released

ILLUSTRATIONS BY MATTHEW HAMS

guidelines that recommend higher doses than the IOM did².

Why, instead of clearing confusion as was the IOM's goal, has the report sown division and unrest? "The IOM was too definitive in its recommendations," says Michael Holick, an endocrinologist at Boston University School of Medicine in Massachusetts, and an outspoken critic of the IOM panel's conclusions. "Basically, the vitamin-D recommendations are based on low-quality evidence," says Gordon Guyatt, a clinician researcher at McMaster University in Hamilton, Ontario, who has been a consultant on various guidelines. "I think admitting that would have made some of the angst disappear."

Poor data is one reason that the panel did not recommend higher doses, say interested observers. "We are not free to just accept enthusiastic reports, unless they are based on comprehensive, well-characterized data sets," says Paul Coates, director of the Office of Dietary Supplements at the National Institutes of Health in Bethesda, Maryland, one of the agencies that had requested the IOM's evaluation. "I think everyone wants to do the right thing, but I would say that the government is inherently more conservative." The episode demonstrates the difficulty of producing public-health advice from disparate and sometimes feeble evidence. The former panel members have been touring the United States and Europe to defend and explain their methods. That's what brought Rosen to the bone club. "In the past 50 years of IOM reports, this one has received the most visibility — fortunately or unfortunately," he tells the audience.

SYSTEMATIC REVIEW

Vitamin D's role in promoting bone health through the regulation of calcium is fairly unassailable, but in the past several years, the medical and scientific communities have become preoccupied with how it might prevent chronic disease. Some physicians recommend supplementation of up to 6,000 international units (IU) a day to make up for the time that people spend indoors. This is less than the amount a fair-skinned person without sun-block might make in half an hour of exposure to the midday summer Sun.

In August 2008, the US and Canadian governments asked the IOM for unbiased advice about how much vitamin D and calcium people need. Rosen and 13 colleagues who were selected to serve on the panel amassed about 1,000 studies on metabolism, vitamin intake and impact on human health. They then ranked the studies by the quality of design and execution. Randomized placebo-controlled studies earned the highest rating. Of roughly 70 such trials, most assessed the effect of vitamin D on falls, fractures and bone quality. About a dozen looked at cancer, cardiovascular disease and diabetes, but because of the way those outcomes were assessed, the panel didn't

RAISING THE STAKES

Sales of vitamin D in the United States have risen dramatically in the past decade.



place much trust in the results. Most of the other research has been observational. A 2008 study, for example, reported that men with low levels of vitamin D were more likely to have heart attacks than were those with higher levels³, but it couldn't rule out other explanations for the link.

The panel met 8 times over 20 months. Its efforts culminated in a 1,132-page report¹ concluding that people should aim for blood levels of 50 nanomoles per litre (nmol/L). This level, the IOM said, can be achieved with 600 IU of vitamin D per day (800 for those older than 70) — an amount that doesn't necessarily require supplementation, because many people would get this naturally from Sun exposure, fatty fish and fortified foods. The IOM also stated that reports of widespread deficiency have been exaggerated; the majority of North Americans already have enough vitamin D; and too much of it could be harmful.

Passions ignited immediately. Physicians and alternative-medicine advocates posted websites and Facebook pages declaring the

that this might encourage the German government, which does not fortify food, to reconsider the issue.

The IOM was interested in a study Amling had published, in which he had measured bone quality and blood levels of vitamin D in the bodies of 675 people who had died in good health (for example, in car accidents and suicides)⁴. Amling concluded that an ideal level for the general population would be 75 nmol/L because everyone above that level had strong bones, and they therefore weren't at a high risk of fractures.

The IOM's mandate was to set the levels that protect most people, but not all. It found that Amling's data supported a 50 nmol/L threshold (which had been suggested elsewhere in the literature) because at that level, only 1% of people in the study had weak bones. But Amling says that the IOM made a mathematical mistake: it should have looked at the risk of weak bones in people at or above a certain level, not in the whole population (see 'Denominator dispute'). Instead of dividing

"PEOPLE TELL ME I DON'T KNOW WHAT I'M DOING. IT HAS BECOME PERSONAL."

IOM guidelines flawed. Some claimed that its recommendations were an industry-motivated scheme to keep people in need of prescription drugs and other costly treatments, a theory that Rosen dismisses. And although conspiracy theories tend to be thin on logic or factual substance, scientists and clinicians have raised some legitimate concerns.

STICKY STATS

Michael Amling, a bone expert at the University Medical Center Hamburg-Eppendorf in Germany is one such critic. He says he was thrilled when Rosen e-mailed him in 2009 to enquire about some of his data. "I wanted to support the work of the IOM," says Amling. He assumed its analysis would conclude that most people were vitamin-D deficient, and

the 7 people with weak bones and levels above 50 nmol/L by all 675 people in the study, he says it should have divided 7 by the 82 individuals with levels above 50 nmol/L. Charles McCulloch, a biostatistician at the University of California, San Francisco, who has no vested interest in vitamin-D thresholds, agrees: the panel should have found that 8.5% of the population above 50 nmol/L had weak bones, and therefore according to its goal of allowing no more than 2.5% of the population to be at risk, Amling's data would support a higher level. "I'm very shocked they made such a basic mathematical mistake," Amling says.

Another researcher whose work received a fair share of the IOM's attention is Heike Bischoff-Ferrari, director of the centre for ageing and mobility at the University of Zurich

in Switzerland. She published a meta-analysis in 2009 that pooled eight clinical trials testing the ability of vitamin-D supplements to reduce falling in elderly people⁵. In her analysis, participants who took daily doses of 700–1,000 IU fell less often than those taking a placebo. Doses below 700 IU made no difference.

When the IOM panel came to analyse Bischoff-Ferrari's data, it decided to include different studies. It removed a study⁶ showing a benefit from doses higher than 800 IU because the study had focused on groups of about 20 people, which the panel considered too small. And it added a trial⁷ that Bischoff-Ferrari had excluded because it hadn't been double-blinded. Once the IOM swapped trials in Bischoff-Ferrari's meta-analysis, the evidence showed no benefit from supplementation. Needless to say, Bischoff-Ferrari and others disagree with the IOM's decision.

With no universal criteria to identify which studies ought to be included in meta-analyses, it's hard to say which team selected the most appropriate ones. What is clear, however, is that a lack of high-quality primary research makes these decisions difficult and prone to bias.

Another criticism levelled at the report has to do with the IOM's warning that too much vitamin D could cause harm. In the only clinical trial claiming risk, elderly women treated with a single 500,000-IU dose of vitamin D annually fell and fractured their bones more often than those in the placebo group⁸. Many researchers find the study ridiculous. "No one absorbs 500,000 IU a day from the Sun, so why would you give that as a supplemental dose?" says Edward Giovannucci, a nutritional epidemiologist at the Harvard School of Public Health in Boston, Massachusetts.

THE DEFENCE

These are just some of the criticisms that Rosen and other members of the former IOM committee have heard. In response to Amling's charge that the IOM made a mathematical mistake, Rosen maintains that the method the IOM used to calculate 1% risk is standard procedure for dietary recommendations. The group was asking about a natural population with wide variation in vitamin-D levels. He adds that other skeletal studies showed no benefit in increasing the threshold above 50 nmol/L.

With regard to Bischoff-Ferrari's meta-analysis, Rosen stands by the IOM's decision to remove studies with few participants because they are sensitive to random errors; including them, he adds, can exaggerate an erroneous finding.

And JoAnn Manson, an epidemiologist at

Harvard Medical School in Boston, dismisses the notion that the mega-dose trial ought to be ignored: "Within the first three weeks of this trial, when serum levels were at or above 100 nmol/L, there was an increased risk of falls and fractures." This trial contributed to the IOM's anxiety about doses that might raise blood levels to this amount. Specifically, the IOM set an upper dosage limit of 4,000 IU — a number it arrived at by taking calculations from various studies. Such precautions are not unexpected, says Guyatt. Health officials put a high value on avoiding recommendations that could prove dangerous over time, he says.

The Endocrine Society's guidelines, which were based on four years of periodic review

began enrolment for a 5-year, 20,000-person trial to test the effect of supplements on cancer and cardiovascular disease.

Reinhold Vieth, a vitamin-D researcher at the University of Toronto in Canada, calls this demand for huge trials "a cop-out". He says that there is good evidence that higher levels of vitamin D would reduce rates of multiple sclerosis, but a clinical trial to test this would require thousands of people and 30 years. "Saying we need perfect, placebo-controlled trials is denying the plausible evidence we have," Vieth says. "At what point do you offer advice?"

Manson urges caution. "We've seen promising correlations before that turned out to be wrong when tested in clinical trials." She offers the example of β -carotene, which showed promise as a cancer preventative in observational studies but proved dangerous in high-dose trials. "Why are all these lessons of history no longer relevant when it comes to vitamin D?" she asks.

Perhaps IOM panel members underestimated the passion present in the vitamin-D field. Physicians who recommend high doses of vitamin D might not want to believe that the evidence they have trusted isn't quite up to par. "One thing I wasn't aware of before, is the tremendous pressure from industry and investigators who are tied to their religious belief in vitamin D," says Rosen.

Guyatt says that much of the current fracas could have been avoided if the IOM panel had been a bit more equivocal in its reporting. But Rosen doesn't regret having been dogmatic

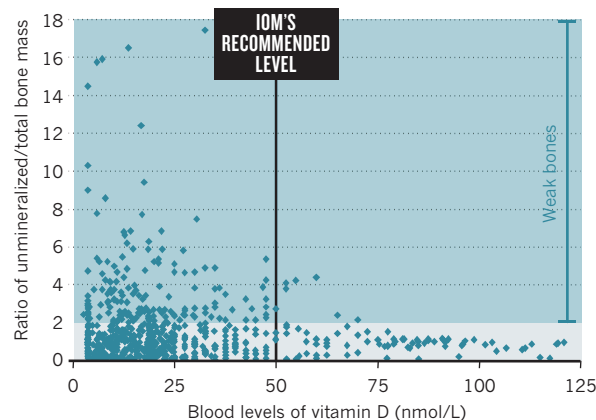
in the recommendations.

Still, he says that he would have liked to take more time to explain the IOM's methods up front rather than just presenting the bottom line. He suspects that unbiased, systematic reviews such as this one will increasingly come under fire when they lead to hard recommendations. "This is the beginning of a whole new phase," he says. "In the old days of medicine we believed experts, and now we say, show us the data." ■

Amy Maxmen is a freelance writer in New York City.

DENOMINATOR DISPUTE

In a study of 675 people (blue dots), 7 people with vitamin-D blood levels of 50 nmol/L or more had weak bones. The Institute of Medicine (IOM) panel concluded that this level met the needs of 99% of the population. But others have disputed the calculation, and recommend a higher level.



of the literature, call people with levels under 50 nmol/L "vitamin-D deficient", and those with levels between 50 nmol/L and 72.5 nmol/L "insufficient". Insufficiency versus deficiency is not a common distinction in guidelines, says Holick, but it reflects the opinion that people benefit from the higher threshold.

The society's guidelines also offer an 'ideal' level of 100–150 nmol/L for non-skeletal health benefits, which would require 1,500–2,000 IU daily, and it advises physicians to monitor vitamin-D levels in healthy people. Quest Diagnostics, a medical-testing corporation headquartered in Madison, New Jersey, that Holick advises, has already begun to implement these deficiency and insufficiency standards over the IOM's, and many physicians are expected to follow suit.

The now-disbanded IOM panel has been formulating a response to clarify how physicians and the public should make sense of the discrepancies in recommendations. The panel's members acknowledge that the case for vitamin D benefiting general health is not closed, and say that the best way to clarify it is with large clinical trials. Manson recently

1. Institute of Medicine *Dietary Reference Intakes for Calcium and Vitamin D* (National Academies Press, 2010).
2. Holick, M. F. et al. *J. Clin. Endocrinol. Metab.* doi:10.1210/jc.2011-0385 (2011).
3. Giovannucci, E., Liu, Y., Hollis, B. W. & Rimm, E. B. *Arch. Int. Med.* **168**, 1174–1180 (2008).
4. Priemel, M. et al. *J. Bone Min. Res.* **25**, 305–312 (2010).
5. Bischoff-Ferrari, H. A. et al. *Br. Med. J.* **339**, b3692 (2009).
6. Broe, K. E. et al. *J. Am. Geriatr. Soc.* **55**, 234–239 (2007).
7. Law, M., Withers, H., Morris, J. & Anderson, F. *Age Ageing* **35**, 482–486 (2006).
8. Sanders, K. M. *JAMA* **303**, 1815–1822 (2010).

COMMENT



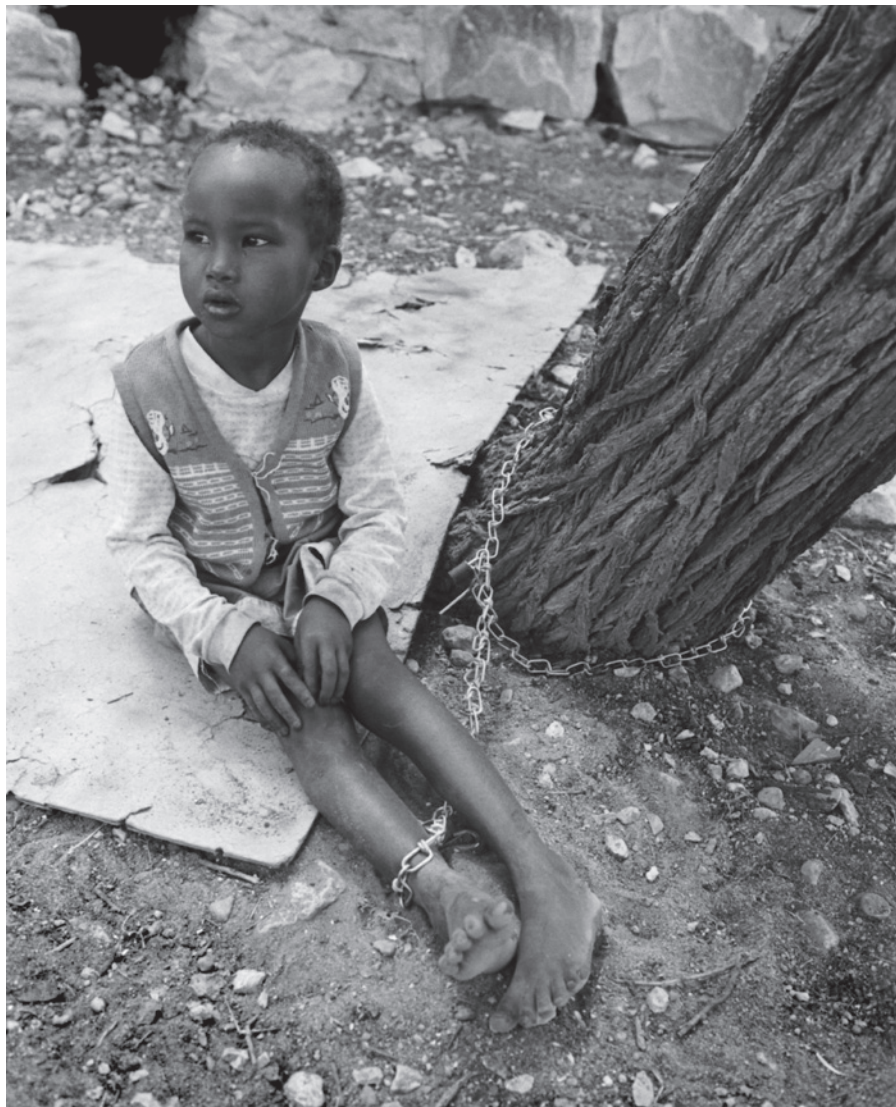
NANOTECHNOLOGY Materials should not be regulated on size alone **p.31**

SUMMER BOOKS Reviewers and editors suggest reading for your holiday **p.32**

CONSERVATION Concern about alien species is scientific and practical **p.36**

EQUALITY Action needed to stop science prizes going primarily to men **p.37**

H. TIMMERMANS/GLOBAL INITIATIVE ON PSYCHIATRY



Improving treatment for children with mental illness, like this girl in Somalia, is an urgent priority.

Grand challenges in global mental health

A consortium of researchers, advocates and clinicians announces here research priorities for improving the lives of people with mental illness around the world, and calls for urgent action and investment.

Schizophrenia, depression, epilepsy, dementia, alcohol dependence and other mental, neurological and substance-use (MNS) disorders constitute 13% of the global burden of disease (Table 1), surpassing both cardiovascular disease and cancer¹. Depression is the third leading contributor to the global disease burden, and alcohol and illicit drug use account for more than 5% (ref. 2). Every seven seconds, someone develops dementia³, costing the world up to US\$609 billion in 2009 (ref. 4). By 2020, an estimated 1.5 million people will die each year by suicide, and between 15 and 30 million will make the attempt⁵.

The absence of cures, and the dearth of preventive interventions for MNS disorders, in part reflects a limited understanding of the brain and its molecular and cellular mechanisms. Where there are effective treatments, they are frequently not available to those in greatest need. In 83% of low-income countries, there are no anti-Parkinsonian treatments in primary care; in 25% there are no anti-epileptic drugs⁶. Unequal distribution of human resources — between and within countries — further weakens access: the World Health Organization's European region has 200 times as many psychiatrists as in Africa⁷. Across all countries, investment in fundamental research into preventing and treating MNS disorders is disproportionately low relative to the disease burden⁸.

To address this state of affairs, the Grand Challenges in Global Mental Health initiative has identified priorities for research in the next 10 years that will make an impact on the lives of people living with MNS disorders. The study was funded by the US National Institute of Mental Health (NIMH) in Bethesda, Maryland, supported by the Global Alliance for Chronic Diseases (GACD), headquartered in London. Answers to the questions posed will require a surge in discovery and delivery science. We use the term 'mental health' as a convenient label for MNS disorders. We exclude conditions with a vascular or infectious aetiology (such as stroke or cerebral malaria), because these fell within the scope of the two previous grand challenges initiatives — in global health and in chronic non-communicable diseases⁹.

This initiative differs from previous priority-setting exercises for mental ►

TABLE 1 | GLOBAL BURDEN OF MENTAL, NEUROLOGICAL AND SUBSTANCE-USE (MNS) DISORDERS*

Rank	Worldwide		High-income countries†		Low- and middle-income countries	
	Cause	DALYs‡ (millions)	Cause	DALYs (millions)	Cause	DALYs (millions)
1	Unipolar depressive disorders	65.5	Unipolar depressive disorders	10.0	Unipolar depressive disorders	55.5
2	Alcohol-use disorders	23.7	Alzheimer's and other dementias	4.4	Alcohol-use disorders	19.5
3	Schizophrenia	16.8	Alcohol-use disorders	4.2	Schizophrenia	15.2
4	Bipolar affective disorder	14.4	Drug-use disorders	1.9	Bipolar affective disorder	12.9
5	Alzheimer's and other dementias	11.2	Schizophrenia	1.6	Epilepsy	7.3
6	Drug-use disorders	8.4	Bipolar affective disorder	1.5	Alzheimer's and other dementias	6.8
7	Epilepsy	7.9	Migraine	1.4	Drug-use disorders	6.5
8	Migraine	7.8	Panic disorder	0.8	Migraine	6.3
9	Panic disorder	7.0	Insomnia (primary)	0.8	Panic disorder	6.2
10	Obsessive-compulsive disorder	5.1	Parkinson's disease	0.7	Obsessive-compulsive disorder	4.5
11	Insomnia (primary)	3.6	Obsessive-compulsive disorder	0.6	Post-traumatic stress disorder	3.0
12	Post-traumatic stress disorder	3.5	Epilepsy	0.5	Insomnia (primary)	2.9
13	Parkinson's disease	1.7	Post-traumatic stress disorder	0.5	Multiple sclerosis	1.2
14	Multiple sclerosis	1.5	Multiple sclerosis	0.3	Parkinson's disease	1.0

*Data from ref. 1. Examples of MNS disorders under the purview of the Grand Challenges in Global Mental Health initiative.

†World Bank criteria for income (2009 gross national income (GNI) per capita): low income is US\$995 equivalent or less; middle income is \$996–12,195; high income is \$12,196 or more.

‡A disability-adjusted life year (DALY) is a unit for measuring the amount of health lost because of a disease or injury. It is calculated as the present value of the future years of disability-free life that are lost as a result of the premature deaths or disability occurring in a particular year.

► health^{10–12} in four ways. First, its scope is global. Second, it is the first to employ the Delphi method¹³, a structured technique using controlled feedback to arrive at consensus within a dispersed panel of many participants. Third, it covers the full range of MNS disorders. Finally, the effort hopes to build a wide-ranging community of research funders — much as the challenge for non-communicable diseases led to the creation of the GACD.

SETTING PRIORITIES

The prioritization exercise assembled the largest international Delphi panel so far on the subject. An executive committee of leaders of key funding agencies provided broad oversight. A scientific advisory board comprising leaders in the relevant scientific disciplines guided the process. And an administrative team from the NIMH worked with the chairs of the committee and advisory board to coordinate communication and data analysis.

The advisory board nominated 594 researchers, advocates, programme implementers and clinicians; 422, working in more than 60 countries, agreed to participate. Researchers in genetics and genomics, neuroscience, basic behavioural science and neurodevelopment made up just over one-third of the panel. Mental-health services researchers constituted another quarter, and a further third were clinical researchers and epidemiologists (see Supplementary Figs 1–3).

In Round 1, panel members were asked to respond to the question “What are the grand challenges in global mental health?” by

listing up to five areas they considered to be top priorities. As in previous initiatives^{14,15}, a ‘grand challenge’ was defined as “a specific barrier that, if removed, would help to solve an important health problem. If successfully implemented, the intervention(s) it could lead to would have a high likelihood of feasibility for scaling up and impact.” Round 1 yielded 1,565 challenges.

These were distilled by the administrative team and chairs of the executive committee and advisory board into a shorter list of 154 unique challenges from which panellists selected their top 40 in Round 2 (Supplementary Table 1). The top 25 challenges from this list are shown in Table 2. Round 3 asked panellists to rank each challenge on a four-point scale for: ability to reduce disease burden; impact on equity; immediacy of impact; and feasibility (see Supplementary Methods). To arrive at the final ranked master list of grand challenges, individual rankings for each challenge were weighted, summed across all four criteria, and divided by the total number of responses (Supplementary Table 3).

THE LIST

The 25 grand challenges in Table 2 run the research gamut from preclinical questions into the aetiology and treatment of MNS disorders, to implementation and policy needs to scale up effective interventions. All the challenges emphasize the need for global cooperation in the conduct of research to create shared access to data, expertise and capacity-building opportunities. Children emerge as requiring particular attention for prevention and care. Most mental disorders

involve developmental processes, so reducing the duration of untreated illness by focusing resources on the earliest definable clinical stage of illness could revolutionize treatment. Similarly, it is imperative that we explore the role of prenatal exposures and develop interventions to reduce the long-term negative impact of low childhood socioeconomic status on cognitive ability.

The challenges capture several broad themes. First, the results underscore the need for research that uses a life-course approach. This approach acknowledges that many MNS disorders either begin or manifest in early life, and is equally attentive to risk factors and disorders affecting children and the elderly. Efforts to build mental capital — the cognitive and emotional resources that influence how well an individual is able to contribute to society and experience a high quality of life — could also mitigate the risk of disorders such as depression, substance-use disorders, bipolar disorder and dementia¹⁶.

Second, the challenges recognize that the suffering caused by MNS disorders extends beyond the patient to family members and communities. Thus, health-system-wide changes are crucial, together with attention to social exclusion and discrimination. At the same time, research into systems interventions, such as integrating care for MNS disorders into chronic-disease care, could transform health services and reduce costs.

Third, the challenges underline the fact that all care and treatment interventions — psychosocial or pharmacological, simple or complex — should have an evidence base to provide programme planners, clinicians and

TABLE 2 | GRAND CHALLENGES FOR MNS DISORDERS

	Top 25 challenges*	Illustrative research questions
Goal A Identify root causes, risk and protective factors	<ul style="list-style-type: none"> Identify modifiable social and biological risk factors across the life course Understand the impact of poverty, violence, war, migration and disaster Identify biomarkers 	<ul style="list-style-type: none"> What is the relationship between early fetal and child development and the onset of MNS disorders? What are the phenotypes and endophenotypes of MNS disorders across cultural settings? What gene–environment interactions are associated with the increased risk for mental disorders? What factors promote resilience and prevent mental disorders in persons at extreme social disadvantage? What role does social context play in the persistence of MNS disorders throughout life?
Goal B Advance prevention and implementation of early interventions	<ul style="list-style-type: none"> Support community environments that promote physical and mental well-being throughout life Reduce the duration of untreated illness by developing culturally-sensitive early interventions across settings Develop interventions to reduce the long-term negative impact of low childhood socioeconomic status on cognitive ability and mental health Develop an evidence-based set of primary prevention interventions for a range of MNS disorders Develop locally appropriate strategies to eliminate childhood abuse and enhance child protection 	<ul style="list-style-type: none"> Which behavioral skills can enhance executive function, resilience and cognitive flexibility throughout life? What neuroprotective agents and/or cognitive retraining paradigms can be used during the period of rapid brain development to reduce vulnerability to disorders in adolescence? How effective are home- and school-based interventions for child abuse and neglect?
Goal C Improve treatments and expand access to care	<ul style="list-style-type: none"> Integrate screening and core packages of services into routine primary health care Reduce the cost and improve the supply of effective medications Develop effective treatments for use by non-specialists, including lay health workers with minimal training Incorporate functional impairment and disability into assessment Provide effective and affordable community-based care and rehabilitation Improve children's access to evidence-based care by trained health providers in low- and middle-income countries Develop mobile and IT technologies (such as telemedicine) to increase access to evidence-based care 	<ul style="list-style-type: none"> How effective are brief screening tools for the detection of MNS disorders in routine care settings? How effective are interventions for serious mental disorders delivered by lay health workers? How will increased understanding of neural circuits lead to alternatives to current pharmacological interventions? How can mobile-phone technology be used to monitor seizure frequency? How can video games and other electronic media be used for cognitive remediation across cultural settings? What psychosocial interventions produce the best outcomes for community-based care for MNS disorders across cultural settings?
Goal D Raise awareness of the global burden	<ul style="list-style-type: none"> Develop culturally informed methods to eliminate the stigma, discrimination and social exclusion of patients and families across cultural settings Establish cross-national evidence on the cultural, socioeconomic and services factors underlying disparities in incidence, diagnosis, treatment and outcomes Develop valid and reliable definitions, models and measurement tools for quantitative assessment at the individual and population levels for use across cultures and settings Establish shared, standardized global data systems for collecting surveillance data on the prevalence, treatment patterns and availability of human resources and services 	<ul style="list-style-type: none"> What are the components of effective interventions to reduce stigma associated with MNS disorders? What interventions to reduce stigma and discrimination can be targeted to and implemented in health and social service settings in different health-system environments? What is the impact of macroeconomic factors (such as unemployment rates, international trade, national income) on the prevalence of MNS disorders over time? What is the impact of policy initiatives on the coverage of treatment for MNS disorders? What measurement factors contribute to differences in the prevalence of mental disorders across ethnic groups within and between countries?
Goal E Build human resource capacity	<ul style="list-style-type: none"> Increase capacity in low- and middle-income countries by creating regional centers for mental-health research, education, training and practice that incorporate the views and needs of local people Develop sustainable models to train and increase the number of culturally and ethnically diverse lay and specialist providers to deliver evidence-based services Strengthen the mental-health component in the training of all health-care personnel 	<ul style="list-style-type: none"> What is the most effective way to train primary health-care workers to deliver evidence-based care with adequate fidelity to guidelines? What is the comparative effectiveness of care for MNS disorders by different cadres of health-care providers? What are the views of low-income communities in high- and low-income countries on the priority research questions for MNS disorders?
Goal F Transform health-system and policy responses	<ul style="list-style-type: none"> Establish and implement minimum health-care standards for MNS disorders around the world Redesign health systems to integrate MNS disorders with other chronic-disease care, and create parity between mental and physical illness in investment into research, training, treatment and prevention Incorporate a mental-health component into international aid and development programmes 	<ul style="list-style-type: none"> What can we learn from different approaches (and associated costs) to integrated delivery of care across health systems? What are the most effective health-system-wide strategies to reduce consumption of alcohol and illicit drugs? What is the impact of legislation that ensures parity between mental and other illnesses on access to mental-health services?
Summary principles	<ul style="list-style-type: none"> Use a life-course approach to study Use system-wide approaches to address suffering 	<ul style="list-style-type: none"> Use evidence-based interventions Understand environmental influences

*The order in which the challenges are presented does not indicate frequency of endorsement or relative importance. **Bold type** denotes the top five challenges ranked by disease-burden reduction, impact on equity, immediacy of impact and feasibility.



Women in Priluki psychiatric hospital, Ukraine.

policy-makers with effective care packages. Finally, the panel's responses underscore important relationships between environmental exposures and MNS disorders. Extreme poverty, war and natural disasters affect large swathes of the world, and we still do not fully understand the mechanisms by which mental disorders might be averted or precipitated in those settings.

NEXT STEPS

There have been some major advances in our understanding of the aetiology and treatment of MNS disorders. Future breakthroughs are likely to depend on discoveries in genomics and neuroscience, in tandem with exploration of the role of sociocultural and environmental contexts. The top five challenges ranked by disease-burden reduction, impact on equity, immediacy of impact, and feasibility should serve as a starting point for immediate research and prioritization of policies (see bold lines in Table 2).

Action on all the challenges will require long-term investment. Substantial research progress can be achieved in the next ten years if funding begins immediately. Already, the NIMH's initiative, Collaborative Hubs for International Research in Mental Health, has committed to support research on the use of non-specialist mental health-care providers and research training in low- and middle-income countries. But a wider set of stakeholders must also be engaged, particularly in problems that require integrated research and policy interventions.

The WHO should disseminate information on these challenges to its member countries' health ministries and research councils to shape research and action priorities. Given

the intimate relationship between economic and social development and the needs of people with MNS disorders, the World Bank, regional development banks, national development agencies, foundations, non-governmental organizations and the global business community should all participate in addressing the challenges.

Researchers and funders have tremendous responsibility in this context. Consortia and networks, advocacy organizations, universities and their partners should organize their activities around one or more of the goals and the attendant grand challenges. The leaderships of the Grand Challenges in Global Mental Health, the GACD and their partners will meet in October to develop a strategy for regular monitoring of progress.

Even incremental progress in addressing the grand challenges in global mental health could lead to significant economic and quality-of-life benefits — including reductions in inappropriate use of health care and increased productivity for years to come¹⁷ — that would far outweigh investment costs. Although the greatest challenge — the elimination of MNS disorders — may not be attainable within the next 10 years, the research suggested above must be conducted forthwith. ■

Pamela Y. Collins *Office for Research on Disparities and Global Mental Health, National Institute of Mental Health, Maryland, USA.* **Vikram Patel**, *Centre for Global Mental Health, London School of Hygiene & Tropical Medicine UK, Sangath, Goa, India.* **Sarah S. Joestl**, *Office for Research on Disparities and Global Mental Health, National Institute of Mental Health,*

USA. **Dana March**, *Office for Research on Disparities and Global Mental Health, National Institute of Mental Health, USA.* **Thomas R. Insel**, *National Institute of Mental Health, USA.* **Abdallah S. Daar**, *University of Toronto and McLaughlin-Rotman Centre for Global Health, Toronto, Canada, and Chair, Global Alliance for Chronic Diseases.*

On behalf of the Scientific Advisory Board and the Executive Committee of the Grand Challenges on Global Mental Health.

e-mails: pamela.collins@nih.gov; a.daar@utoronto.ca

1. World Health Organization *The Global Burden of Disease: 2004 Update* (WHO, 2008).
2. WHO *Atlas on Substance Use* (WHO, 2010).
3. Ferri, C. P. et al. *Lancet* **366**, 2112–2117 (2005).
4. Wimo, A., Winblad, B. & Jönsson, L. *Alzheimer's & Dementia* **6**, 98–103 (2010).
5. Bertolote, J. & Flieschmann, A. *Suicidologia* **7**, 6–8 (2002).
6. WHO *Country Resources for Neurological Disorders 2004* (WHO, 2004).
7. WHO *Mental Health Atlas* (WHO, 2005).
8. Saxena, S., Thornicroft, G., Knapp, M. & Whiteford, H. *Lancet* **370**, 878–889 (2007).
9. Daar, A. S. et al. *Nature* **450**, 494–496 (2007).
10. Lancet Mental Health Group *Lancet* **370**, 1241–1252 (2007).
11. Sharan, P. et al. *Br. J. Psychiatry* **195**, 354–363 (2009).
12. Tomlinson, M. et al. *Bull. WHO* **87**, 438–446 (2009).
13. Jones, J. & Hunter, D. *Br. Med. J.* **311**, 376–380 (1995).
14. Daar, A. S. et al. *Nature* **450**, 494–496 (2007).
15. Varmus, H. et al. *Science* **302**, 398–399 (2003).
16. Beddington, J. et al. *Nature* **455**, 1057–1060 (2008).
17. Rupp, A. *Br. J. Psychiatry* **166**, 26–33 (1995).

Supplementary Information and a full list of authors accompany this article online at www.nature.com/nature and can also be viewed at <http://grandchallengesgmh.nimh.nih.gov>.

Don't define nanomaterials

Basing regulations on a term with no scientific justification will do more harm than good, argues **Andrew D. Maynard**.

Five years ago, I was a proponent of a regulatory definition of engineered nanomaterials. I have changed my mind. With policy-makers looking for clear definitions on which to build 'nano-regulations', there is a growing danger of science being pushed aside.

It makes sense to assume that nanomaterials could come with unanticipated risks. A rapidly growing body of research indicates that some nanoscale materials behave differently from their bigger and smaller counterparts¹. For instance, normally benign titanium dioxide — widely used as a whitener — becomes increasingly toxic as its particle size shrinks. Nanoscale titanium dioxide has been classified as a potential human carcinogen by the US National Institute for Occupational Safety and Health.

But it is becoming clear that many parameters other than size modulate risk, including particle shape, porosity, surface area and chemistry. Some of these parameters become more relevant at smaller scales — but not always. The transition from 'conventional' to 'unconventional' behaviour, when it does occur, depends critically on the particular material and the context.

A 'one size fits all' definition of nanomaterials will fail to capture what is important for addressing risk.

definition of engineered nanomaterials is needed to ensure their safe use under such regulations. But a sensible definition has proved hard, if not impossible, to arrive at.

Last year, the European Commission (EC) released this draft definition for public comment: "a material that consists of particles with one or more external dimensions in the size range 1 nm–100 nm for more than 1% of their number"; and/or "has internal or surface structures in one or more dimensions in the size range 1 nm–100 nm"; and/or

To avoid similar problems with nanomaterials, regulators should work with a list of nine or ten attributes (including size and surface area) for which certain values trigger action. This list of trigger points would differ from the EC black-and-white definition in several key ways: attributes other than size and surface area would be included; cut-off values would be compound-specific and based on current science; and the criteria would allow for changes in a given material over time. Such regulatory sophistication will obviously

take a lot of work. Much more research is needed to pin down how much a material would need to change to trigger a regulatory red flag: should a 1% change in mean particle size of a material previously determined to be safe, say, raise concerns, or should it be a 50% change? But enough is known today for an expert panel to begin determining key attributes and preliminary trigger points for many materials.

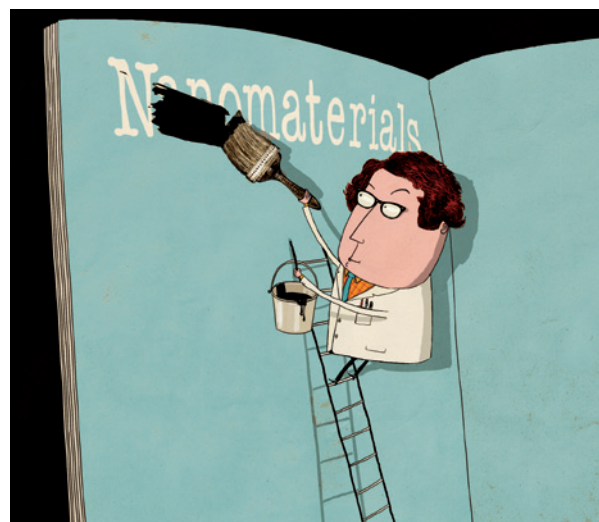
These trigger points must be flexible, so that they can be modified as evidence grows. Adaptive regulations are necessary to respond to scientific evidence, although current US legal mechanisms make them hard to enact.

Last month, the US Executive Office of the President released policy principles that stress the need for science-based regulation of engineered nanomaterials, stating: "A focus on novel properties and phenomena observed in nanomaterials may ultimately be more useful than a categorical definition based on size alone."² This is an important step in the right direction. Even so, assuming that nanomaterials are a unique class of material continues to present a stumbling block to effective regulation. ■

Andrew D. Maynard is director of the Risk Science Center at the University of Michigan School of Public Health, Ann Arbor, Michigan 48109-2029, USA.
e-mail: maynarda@umich.edu

1. Maynard, A. D., Warheit, D. & Philbert, M. A. *Toxicol. Sci.* **120**, S109–S129 (2011).
2. Holdren, J. P., Sunstein, C. R. & Siddiqui, I. A. *Policy Principles for the U.S. Decision-making Concerning Regulation and Oversight of Applications of Nanotechnology and Nanomaterials* (Executive Office of the President, 2011).

Further reading accompanies this article online at go.nature.com/xrp2jp



TERM OF ART OR SCIENCE?

Concern over the risks of engineered nanomaterials has led organizations around the world to reconsider their regulations. In June, the US Food and Drug Administration issued draft guidance to help companies determine whether their products utilize nanomaterials. The US Environmental Protection Agency has just issued draft guidelines for when a nanomaterial should be considered a 'new' substance under the Federal Insecticide, Fungicide and Rodenticide Act, even if a larger-scale version is already in use.

In Europe, the REACH regulations (on the registration, evaluation, authorization and restriction of chemicals), in force since 2007, have been accompanied by contentious (and so far unresolved) discussions about how they apply to engineered nanomaterials. New European cosmetics regulations require ingredients that are engineered nanomaterials to be listed on product labels by 2013.

Many argue, and I once agreed, that a

"has a specific surface area by volume greater than $60 \text{ m}^2 \text{ cm}^{-3}$, excluding materials consisting of particles with a size lower than 1 nm". Despite a stated intent to base the definition on "available scientific knowledge", none of these criteria is supported by current data on nanomaterial risk.

Unsurprisingly, the commission has not yet been able to reach a consensus on a definition. The coordinator of the EC's nano team has been quoted as saying that ultimately the decision on a regulatory definition of nanomaterials would be a "policy one". This should ring alarm bells throughout the scientific community: it implies that the basis for nanomaterial regulation will be a 'term of art' rather than science.

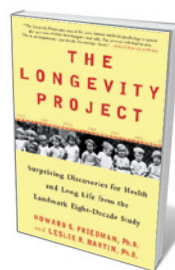
Strict definitions of harmful materials have caused problems before. The mineral Libby vermiculite, for instance, contains deadly asbestiform fibres — and yet it slipped through the regulatory net for many years because it didn't fit the official definition of asbestos.



SUMMER BOOKS

As holidays beckon, *Nature's* reviewers and editors offer a selection of reading for researchers away from the bench and lecture hall.

EDITORS' PICKS



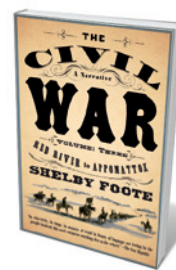
The Longevity Project

HOWARD S. FRIEDMAN AND
LESLIE R. MARTIN
*Hudson Street Press/Hay Hack
House: 2011.*

In 1921, a remarkable study began tracking the lives and loves of 1,528 US citizens from childhood to death. The study continues today, with research teams — led by Howard Friedman, a psychology professor at the University of California, Riverside — re-examining the data and keeping tabs on the few subjects who are still alive. With lucid prose and rigorous analysis, Friedman and his colleague Leslie Martin chronicle and interpret the findings from the 90-year project (*Nature* 471, 443–444; 2011).

Analysing copious data, they establish what it is about the participants' personalities, life circumstances and behaviours that led some to stay healthy and others to fall ill or die before their time. Is longevity associated with being married, daily jogs, living with pets or faith in God? Is it more salubrious to retire early to a serene tropical island or to persevere in a stressful but challenging occupation? The book offers surprising lessons. Many a tenacious scientist will be gratified to learn, for example, that purpose, planning, prudence and persistence have important roles in maintaining a long life.

Sonja Lyubomirsky is professor of psychology at the University of California, Riverside, USA.



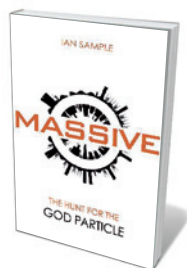
The Civil War: A Narrative

SHELBY FOOTE
*Vintage Books: 1958 (Vol. 1);
1963 (Vol. 2); 1974 (Vol. 3).*

To commemorate the 150th anniversary of the American Civil War, I've been staying up way too late, reading this savagely compelling epic history of the battles fought on the field and in the political arenas of Washington DC and Richmond, Virginia. The war showed the coming power of the Industrial Revolution. Soldiers fought with semi-automatic carbines and iron-clad ships, while railways and the telegraph transformed transport and communications. Yet military success ultimately hinged on old technology: bayonet charges, gruelling marches on foot and paper battle plans that often went astray.

Given my own work on conservation strategy, I found it fascinating to see the interplay between the best-laid plans and the challenges of enacting them in the fog of war. Like the generals, conservationists share a moral cause, incomplete information, limited resources, poorly trained troops and impatient political masters. Our success, too, depends on developing, sharing and implementing good plans, as well as adapting to changing circumstances in the most dire straits. I only hope our heroism matches even a fraction of that shown by the Civil War combatants.

Nick Salafsky is co-director of the conservation non-profit organization Foundations of Success, Bethesda, Maryland, USA.



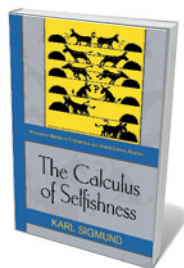
Massive: The Hunt for the God Particle

IAN SAMPLE
Virgin/Basic Books: 2010.

Massive describes the long search for the tiny Higgs boson, or 'God particle', the existence of which would verify the origins of mass and unravel many of the mysteries of the building blocks of the Universe. Although named after Scottish physicist Peter Higgs, the theory behind the boson was pieced together in the 1960s by three groups of scientists simultaneously (including Higgs) in different countries (*Nature* **465**, 873–874; 2010).

Science journalist Ian Sample enlivens the tale through anecdotes, spanning Higgs's initial presentation of his ideas about the particle at Princeton University, New Jersey, in 1966 to the current search for the boson using the Large Hadron Collider (LHC) at CERN, Europe's particle-physics lab in Geneva, Switzerland. He relates how technical problems with the LHC delayed the experiment, and how fears were spread in the media that the creation of exotic 'strangelet' particles could lead the world to collapse in a split second. Although some scientists, including Stephen Hawking, believe that the Higgs boson will never be found, Higgs and others are optimistic that the search is coming to an end.

Sara de Freitas is director of research at the Serious Games Institute at Coventry University, UK.



The Calculus of Selfishness

KARL SIGMUND
Princeton University Press: 2010.

By telling you how to behave so that people will like you, this book will help you become a well-regarded member of your community. 'Give and you shall receive' is a rule not only for religious believers — it is supported by evolutionary theory.

Even when we are acting selfishly, explains evolutionary game theorist Karl Sigmund in his aptly titled *The Calculus of Selfishness*, we gain most by being cooperative (*Nature* **464**, 1280; 2010). It is a win-win strategy. Eighteenth-century economist Adam Smith knew it: "It's not from the benevolence of the butcher, the brewer, or the baker, that we expect our dinner, but from their regard to their own self-interest."

Sigmund's insights are more sophisticated. Because trust is a lubricant of social life, he says, concerns for reputation have a vital role in how we behave. This preoccupation with how people perceive us is hard-wired into our brains. We unconsciously become nicer to others when we feel we are being observed — even a picture of watching eyes is enough to change our behaviour. Sharing information through gossip and paying to punish freeriders are also ways in which nice guys can finish first.

Manfred Milinski is director at the Max-Planck-Institute for Evolutionary Biology, Ploen, Germany.



Virunga: The Survival of Africa's First National Park

EDITED BY MARC LANGUY
AND EMMANUEL DE MERODE
Lannoo: 2009.

Virunga National Park in the Democratic Republic of the Congo is Africa's oldest and most species-rich national park. Its mountain gorillas, of which only 790 remain, are powerful reminders of human evolutionary history. Yet the park is in great danger, caught between population growth, agricultural encroachment, poverty and the devastation of years of civil war. Its survival hinges on a network of dedicated people and organizations, some of who risk their lives: more than 130 rangers have been killed since the mid-1990s.

This book is a scientific account of 80 years of work at Virunga, co-edited by the park's current director, Emmanuel de Merode. It explores the park's natural wealth, its vulnerability and the challenges of safeguarding it. Beautiful images and maps throughout reconnect the reader with the richness that the world has brought forth — ours to preserve or squander. Reading the book left me concerned for Virunga's future, but also invigorated with the gift of being alive.

Wolfgang Lucht is professor of sustainability science at the Potsdam Institute for Climate Impact Research and at Humboldt University Berlin, Germany.



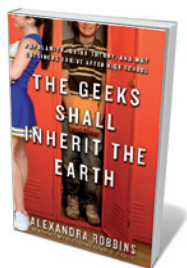
23 Things They Don't Tell You About Capitalism

HA-JOON CHANG
Allen Lane: 2010/Bloomsbury: 2011.

Ha-Joon Chang takes 23 oft-asserted 'truths' about markets, then debunks them in a witty way. Of the assumption that pay and productivity should be linked, he argues that chief executives in the United States cannot have become ten times more productive since the 1960s, even though their salaries have risen from 30–40 times that of the average worker to 300–400 times that today. And he puts paid to views that free markets left unchecked will make us all better off, noting the widening gulf between the super-rich and everyone else.

As the complexity of the world outstrips our ability to control it, he concludes that we need an intelligent state acting in the interests of society as a whole. We need to find a better balance between the finance sector and the real economy in which people make things. From my environmentalist's standpoint, Chang now needs to address two crucial issues: how economics should take into account the unpriced and undervalued natural world of soils, water and atmosphere on which our life depends; and how we can embed long-term thinking in government and corporate decisions.

Camilla Toulmin is director of the International Institute for Environment and Development (IIED), London, UK.



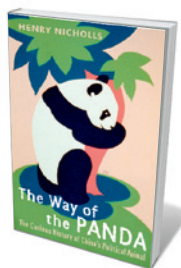
The Geeks Shall Inherit the Earth: Popularity, Quirk Theory, and Why Outsiders Thrive After High School

ALEXANDRA ROBBINS
Hyperion: 2011.

What do Bill Gates, Lady Gaga and J. K. Rowling have in common? They were all outsiders in high school. In this moving chronicle of US teens, journalist Alexandra Robbins proposes that the traits that mark students as different help them to succeed later in life. She followed seven students for a year, including a boy from Hawaii preoccupied with gaming, a Georgia girl concerned with her sexual orientation and a Virginia boy fascinated with IQ challenges. The students' courage and commitment to their own distinctiveness helped them to turn their talents into academic and social achievement.

Writing from the perspective of social psychology, Robbins says little about conditions such as autism spectrum disorders. But her assertion that diversity is more than race or ethnicity puts her in line with the neurodiversity movements. It is increasingly 'cool' to be smart or to have scattered skills in high school.

Roy Richard Grinker is professor of anthropology, international affairs and human sciences at George Washington University, Washington DC, USA.



The Way of the Panda: The Curious History of China's Political Animal

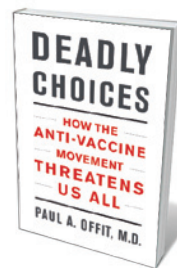
HENRY NICHOLLS
Pegasus/Profile Books: 2011.

There are now vastly more giant-panda products — including films, toys, blogs and books — than there are living giant pandas. Estimates put the number of wild and captive animals at around 3,000. Even with international efforts to save it, the giant panda is an iconic animal for all endangered species on the planet.

For those wishing to learn more about this captivating animal, three books stand out for veracity, clarity and history. The most recent is *The Way of the Panda*, by science writer Henry Nicholls, who splendidly relates the way giant pandas have taken hold of human hearts and politics (*Nature* **468**, 503–504; 2010).

Nicholls generously cites the importance of two books by other naturalists: George B. Schaller's *The Last Panda* (University of Chicago Press, 1993) and *Men and Pandas* by Desmond and Ramona Morris (McGraw-Hill, 1967). Coming soon is *The Giant Book of the Giant Panda* (Smithsonian) by Pan Wenshi, the world's foremost giant-panda expert.

Nancy Lee Nash was a WWF consultant 1979–84 and wrote the successful proposal for China–WWF contact in 1979.



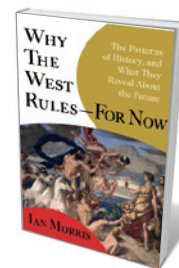
Deadly Choices: How the Anti-Vaccine Movement Threatens Us All

PAUL OFFIT
Basic Books: 2011.

I've been teaching a seminar on pseudoscience that deals with the anti-vaccine movement. We've been looking at Paul Offit's book, which discusses the parents and others whose lack of understanding of, or belief in, science leads to the growth of this movement. It points out the often serious and sometimes deadly consequences of not vaccinating your child. Offit discusses the herd immunity that we have counted on in the past, and how it diminishes as the anti-vaccine movement grows.

My own daughter, living in California, has been affected. Because of the whooping-cough epidemic that killed several babies in the state in the past year, she had to limit her baby's travel for several months until the infant could be vaccinated. When your granddaughter is endangered, it concentrates the mind on the foolish and unnecessary hazard that some have created by not vaccinating their children and recommending the same for others. By the end of the course, my students had a better appreciation of the downside of superstition and pseudoscience. If read with an open mind, Offit's book should convince the rest of the world as well.

Jay Pasachoff is director of the Hopkins Observatory, Williams College, Massachusetts, USA.



Why the West Rules – For Now: The Patterns of History, and What They Reveal About the Future

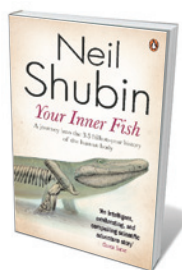
IAN MORRIS
Farrar Straus and Giroux: 2010.

Are the Western civilizations that spread from Egypt, Sumeria, Greece and Rome predestined to get ahead of those in the East? Yes and no, according to historian Ian Morris. The society with the edge is the one that can best organize and direct people, technology and finance into systems that work for the time. This now favours the West, but that hasn't always been true and needn't be so in the future.

Morris answers the question quantitatively. He begins with the DNA-dispersion patterns laid down when people moved out of Africa. Incorporating evidence of worship, urbanization, centralized authority, tools and the harnessing of energy, he analyses many societies, over 100- and 1,000-year intervals. The West usually scores higher than the East, but not always.

He sees the next 40 years as the most crucial in human history: will we be able to manage what has been unleashed by consumption and energy use? Morris contends that less liberal, more centralized approaches will pull ahead. This is a thought-provoking book for science and history buffs alike.

Margaret Catley-Carlson is chair of the Crop Diversity Trust, Rome, Italy.



Your Inner Fish: A Journey Into the 3.5 Billion- Year History of the Human Body

NEIL SHUBIN
Vintage/Allen Lane: 2008.

When you prepare your next salmon for dinner, think about how similar its body plan is to yours, at least relative to other phyla such as corals or oysters. Neil Shubin, palaeontologist and co-discoverer of *Tiktaalik*, the ‘fish-with-hands’ fossil unearthed in the remote Inuit land of northern Canada, tells the intriguing story of the uncovering of new links in the evolutionary chain from fish to amphibians (*Nature* **451**, 245; 2008).

Forget *Ichthyostega*, the famous hybrid fossil you learned about in school; it is practically an amphibian, its fins already evolved into fingers and toes. The earlier stages in the evolution of our wrist, palm and digits are apparent in *Tiktaalik*, implying that the basic skeleton of our hands and feet emerged more than 360 million years ago. From the reptiles that lived 250 million years ago to today, the five-finger hand has been the optimal solution. But imagine what might have been if seven fingers per hand was instead the norm? How would piano concertos for fourteen-finger hands sound? Shubin's fine book raises thoughts such as these.

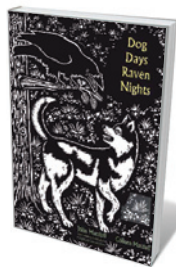
Birger Schmitz is professor of geology at the University of Lund, Sweden.



ANNA LIMES

Dog Days, Raven Nights

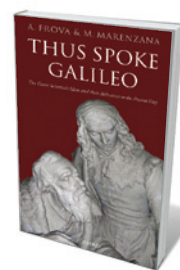
JOHN M. MARZLUFF AND COLEEN MARZLUFF
Yale University Press: 2011.



Ravens are intensely social and famously clever, but elusive to those who would observe them. John and Coleen Marzluff team up with veteran raven expert Bernd Heinrich in the Maine woods to study an important aspect of raven social behaviour: why they recruit other ravens to a food source. The picture that emerges is far from a simple tale of instinct and reflexes. Social life among ravens, like that of primates, is intricate, flexible and subtle.

Behind the scenes, Sitka, a Siberian husky, and Topper, a black labrador, are working dogs who pull dead calves on a sled to feed the ravens in the aviary. Husky pups are trained to pull sleds, and are mentored by the experienced Sitka. In a race, these spirited animals maintain a pace of more than 5 kilometres a minute, running the equivalent of five marathons a day for ten days. Full of the grittiness of experimental persistence — and the splendour of ravens and dogs — this is a warm tale of wonderful science.

Patricia Churchland is professor of philosophy at the University of California, San Diego, USA.



Thus Spoke Galileo: The Great Scientist's Ideas and Their Relevance to the Present Day

ANDREA FROVA AND
MARIAPIERA MARENZANA
Oxford University Press: 2011.

Most people have heard of Galileo Galilei, but few have read his words. This book lets him speak, conveying his message of reason, intellectual honesty and free thinking, which is of great relevance today. It made me realize that Galileo still influences the modern world; for instance, in defending the supremacy of rationality above dogmatic beliefs. As he argued, we should not accept from science only what is impossible to refute, but we should value the concept of a mutable truth and the use of reasoning in general.

At a time when irrationality is rife in society, we should incorporate these basic tenets of the Galilean enterprise into the education of young people. This anthology, by physicist Andrea Frova and literature scholar Mariapiera Marenzana, spans the scientific, historical and literary viewpoints of Galileo's arguments. Even the errors made by the great Tuscan scientist are a source of further understanding. Galileo can still teach us how to appreciate the marvellous and simple ways in which nature operates.

Bruno Scrosati is professor of electrochemistry at the University of Rome Sapienza, Italy.

CORRESPONDENCE

Non-natives: 141 scientists object

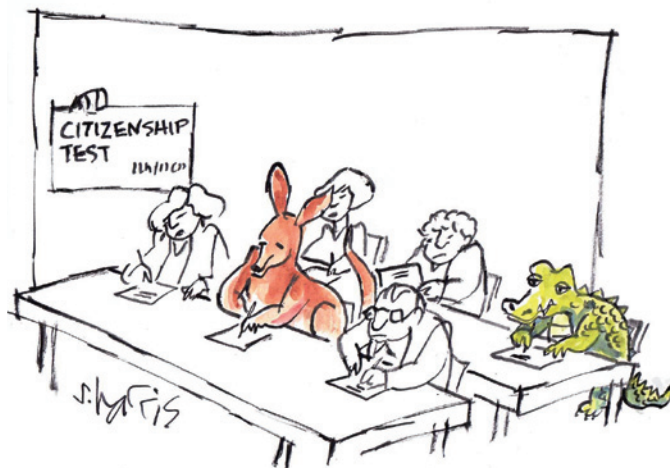
We the undersigned feel that in advocating a change in the environmental management of introduced species (*Nature* 474, 153–154; 2011), Mark Davis and colleagues assail two straw men.

First, most conservation biologists and ecologists do not oppose non-native species per se — only those targeted by the Convention on Biological Diversity as threatening “ecosystems, habitats or species”. There is no campaign against all introductions: scarcity of resources forces managers to prioritize according to the impact of troublesome species, as in the Australian Weed Risk Assessment.

Second, invasion biologists and managers do not ignore the benefits of introduced species. They recognize that many non-native species curtail erosion and provide food, timber and other services. Nobody tries to eradicate wheat, for instance. Useful non-native species may sometimes still need to be managed because they have a negative impact, such as tree invasions that cause water loss in the South African fynbos.

Davis and colleagues downplay the severe impact of non-native species that may not manifest for decades after their introduction — as occurred with the Brazilian pepper shrub (*Schinus terebinthifolius*) in Florida (J. J. Ewel in *Ecology of Biological Invasions of North America and Hawaii* (eds H. A. Mooney and J. A. Drake) 214–230; Springer, 1986). Also, some species may have only a subtle immediate impact but affect entire ecosystems, for example through their effect on soils.

Pronouncing a newly introduced species as harmless



can lead to bad decisions about its management. A species added to a plant community that has no evolutionary experience of that organism should be carefully watched.

For some introductions, eradication is possible. For example, 27 invasive species have been eradicated from the Galapagos Islands, mitigating severe adverse effects on endemic species. Harmful invasive species have been successfully kept in check by biological, chemical and mechanical means.

The public must be vigilant of introductions and continue to support the many successful management efforts.

Daniel Simberloff* *University of Tennessee, Tennessee, USA.*
dsimberloff@utk.edu

**On behalf of 141 signatories (see go.nature.com/fl1eqjn).*

Non-natives: put biodiversity at risk

Bias against non-native species is not xenophobic (*Nature* 474, 153–154; 2011) — it has a sound scientific foundation.

The non-native status of a species is highly relevant to assessing its potential environmental and economic

impact. Unrestrained growth and environmental damage follow when there are no natural enemies in newly colonized areas. This is not necessarily a sign of an invader's superior evolutionary fitness: it may lead to a population collapse due to overexploitation of resources.

Non-native species can increase the variety of species in a community, but it is an oversimplification to equate this with increased biodiversity, of which species richness is only one component. Surviving populations of native species may shrink or become restricted to poor-quality marginal habitats. Such unevenness hardly contributes to a more diverse community.

The genetic diversity of invaded communities may decrease because of bottlenecks: native genotypes disappear as populations fall, whereas the invaders originate from very few initial colonizers.

Establishment of non-native species inevitably decreases global diversity. Australia, for example, was unique in having no placental mammals; their introduction by humans made the continent ecologically more similar to the rest of the world.

Andrei Alyokhin *University of Maine, Maine, USA.*
andrei.al yokhin@umit.maine.edu

Non-natives: plusses of invasion ecology

Contrary to the implications of Mark Davis and colleagues (*Nature* 474, 153–154; 2011), invasion ecology has given us valuable insight into the effects of new species on ecological function and into some of the precipitous changes we may face in the coming decades.

Invasion ecologists generally assert that only a very small fraction of non-native species harm their new ecosystems. This position emerged as early as 1986 and was mainstream in the era that Davis and colleagues claim as the nadir of ecological nativism.

It is unfair to characterize any scientific discipline solely by past failures and to ignore its successes. Invasion ecology is making real progress with defining impact and characterizing risk. Let's not throw up our hands in despair just yet.

Julie L. Lockwood *Rutgers, The State University of New Jersey, USA.*

lockwood@aesop.rutgers.edu

Martha F. Hoopes *Mount Holyoke College, Massachusetts, USA.*

Michael P. Marchetti *California State University, California, USA.*

Non-natives: four risk factors

Mark Davis *et al.* set an unrealistically high bar for those making management decisions about exotic species (*Nature* 474, 153–154; 2011). Control is often easier, cheaper and more effective soon after detection (R. A. Haack *et al. Annu. Rev. Entomol.* 55, 521–546; 2010). We agree that research on ecosystem impact is necessary, but such studies can take years. Meanwhile, we suggest that

control priorities for potential invasive species could be based on easily available data about natural history and evolutionary ecology. We propose four guidelines for identifying such invasives.

An exotic organism may be more likely to invade and cause disruption the greater its rate of reproduction; the greater its dispersal ability; the closer (phylogenetically) its preferred food in its native range is to an abundant taxon in the new range; and the farther away (phylogenetically) its predators and pathogens are in its native range from those in its new range.

For example, the red turpentine beetle (*Dendroctonus valens*) is not particularly disruptive in its native range in North America because it attacks only trees that are already weakened. In China it attacks and kills healthy trees (Z. Yan *et al. Biodivers. Conserv.* **14**, 1735–1760; 2005). The reasons for this beetle's success as an invasive include its high dispersal and reproductive rates, its affinity for Chinese pines closely related to those it feeds on 'at home', and the lack of predators or pathogens phylogenetically similar to ones found in North America.

Manuel Lerdau *Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Menglun, Yunnan, China; and University of Virginia, Virginia, USA.*
mlerdau@virginia.edu

Jacob D. Wickham *Institute of Chemistry, Chinese Academy of Sciences, Beijing, China.*

UK cancer genetics gets personal

There is a promising way in which personalized medicine can be incorporated into health-service infrastructure (*Nature* **473**, 253–254; 2011). In the United Kingdom, the charity Cancer Research UK is leading a partnership with AstraZeneca, Pfizer and the government's Technology Strategy Board to help the National Health Service to adopt a consistent approach to

genetic testing for targeted cancer therapies, and to promote further research into personalized treatment.

The first phase will run from 2011 to 2013 and cost £5.5 million (US\$8.8 million). The programme will model the processes required for routine testing of tumour molecular characteristics and for secure storage and retrieval of molecular and clinical data for research. It will involve seven Experimental Cancer Medicine Centres and 9,000 patients with one of six tumour types: breast, colorectal, lung, prostate, ovary and metastatic melanoma. Up to 22 mutations will be tested, with the aim of harmonizing practices across the centres and labs.

The second phase will establish a molecular diagnostics service to deliver high-quality, standardized tests for patients nationwide and to obtain routine consent for the collection, storage and research use of data on genetics, treatments and outcomes. The long-term strategy includes the flexibility to accommodate new technologies, other cancer types and other disease areas.

Cancer Research UK is in discussion with similar initiatives in the United States, France, Australia and elsewhere to exchange information on mistakes and successes.

David Wiseman, Alice Tuff, James Peach *Cancer Research UK, London, UK.*
david.wiseman@cancer.org.uk

Nuclear winter was and is debatable

Alan Robock's contention that there has been no real scientific debate about the 'nuclear winter' concept is itself debatable (*Nature* **473**, 275–276; 2011).

This potential climate disaster, popularized in *Science* in 1983, rested on the output of a one-dimensional model that was later shown to overestimate the smoke a nuclear holocaust might engender. More refined estimates, combined with advanced three-dimensional models (see go.nature.com/kss8te), have dramatically

reduced the extent and severity of the projected cooling.

Despite this, Carl Sagan, who co-authored the 1983 *Science* paper, went so far as to posit "the extinction of *Homo sapiens*" (C. Sagan *Foreign Affairs* **63**, 75–77; 1984). Some regarded this apocalyptic prediction as an exercise in mythology. George Rathjens of the Massachusetts Institute of Technology protested: "Nuclear winter is the worst example of the misrepresentation of science to the public in my memory," (see go.nature.com/yujz84) and climatologist Kerry Emanuel observed that the subject had "become notorious for its lack of scientific integrity" (*Nature* **319**, 259; 1986).

Robock's single-digit fall in temperature is at odds with the subzero (about –25 °C) continental cooling originally projected for a wide spectrum of nuclear wars. Whereas Sagan predicted darkness at noon from a US–Soviet nuclear conflict, Robock projects global sunlight that is several orders of magnitude brighter for a Pakistan–India conflict — literally the difference between night and day. Since 1983, the projected worst-case cooling has fallen from a Siberian deep freeze spanning 11,000 degree-days Celsius (a measure of the severity of winters) to numbers so unseasonably small as to call the very term 'nuclear winter' into question.

Russell Seitz *Massachusetts, USA.*
seitz@physics.harvard.edu

All-male line-up yet again

Most prestigious prizes in science that are not set aside for women go primarily or only to men. The eight male 2010 Kavli prizewinners in astrophysics, nanoscience and neuroscience are the most recent examples (see go.nature.com/5xh17n).

The Kavli winners are accomplished and deserve their honours. But the frequency of all-male line-ups, and the number of meritorious women,

indicates that women as a group are disproportionately overlooked. Why?

Gender schemas — cognitive structures that summarize our beliefs about the sexes — portray women primarily as nurturing and communal, and men as capable of independent action and work-oriented (V. Valian *Why So Slow? The Advancement of Women*; MIT Press, 1998). Such schemas mean that women's names are unlikely to come to nominators' minds; if women are considered, they are less likely than men to be perceived as prizeworthy (see also *Nature* **469**, 472; 2011).

Prizes matter in part because young women with scientific abilities and interests are more likely to aim high if they see examples of women receiving top awards. Why stay in a field where you have little chance of recognition? We are squandering the abilities of half the human race.

Prize committees need to learn where, how and why our perceptions give men an edge. Committees also need actively to solicit nominations of women and members of under-represented groups. Few guidelines, including those for the Kavli prize, include such encouragements. It is time to stop this cycle of neglect of outstanding women in science. **Virginia Valian** *Hunter College and CUNY Graduate Center, New York, USA.*
virginia.valian@hunter.cuny.edu

CORRECTIONS

An editing change (T. Leitner *et al. Nature* **473**, 284; 2011) confused the point that phylogenetic experts should formulate an a priori hypothesis based on HIV epidemiological information.

B. Bosquet (*Nature* **474**, 36; 2011) notes that Cameroon's REDD Project Idea Note was prepared with the WWF's full support. *Nature's* headline was not intended to undermine efforts by governmental and non-governmental organizations.

Plate motion and mantle plumes

A model based on geophysical data from the Indian Ocean suggests that a mantle-plume head may once have coupled the motions of the African and Indian tectonic plates, and determined their respective speeds. [SEE ARTICLE P.47](#)

R. DIETMAR MÜLLER

News reports persistently remind us of the effects that plate tectonics has on our daily lives. Earthquakes and tsunamis are consequences of the relentless grinding of tectonic plates past each other. In Earth's deep mantle, rocks flow like warm toffee, and hot, buoyant material can rise up to the surface, causing volcanic eruptions and ash clouds that can bring air traffic to a halt. Our knowledge of what ultimately drives the plates is still incomplete, however, not least when it comes to understanding the possible influence of mantle plumes. On page 47 of this issue, Cande and Stegman¹ offer a fresh perspective on the possible coupling between plate motions and the forces generated by these plumes.

Mantle plumes are hot, cylindrical upwellings capped by an enlarged, mushroom-like head. When a rising plume head arrives at Earth's surface, volcanoes spew out scorching volcanic ashes, gases and voluminous lava. These are quite rare geological events. But when they happen they can cause the break-up of supercontinents, as well as mass extinctions and major ecological changes on land and in the oceans². On the bright side, they are also known to act as a catalyst for emplacing diamonds in the shallow crust³.

Cande and Stegman¹ extend the catalogue of phenomena attributed to mantle plumes by suggesting that they are able to impart a substantial force on plates — a plume-push force — that can do more than just break up continents. Their hypothesis is that if a rising plume head impinges on the base of a tectonic plate long after supercontinent break-up and dispersal, its pushing force may result in a substantial transient acceleration or deceleration of plates. Whether a plate adopts the speed of a tortoise or a hare after being hit by a plume depends on how the plume-push force balances out with other forces acting on the plate.

Cande and Stegman study the effect of the Réunion plume, whose mushroom head is thought to have reached the surface around 67 million years (Myr) ago, emplacing the huge volumes of volcanic rocks in India known as the Deccan traps⁴. Using a careful analysis of marine geophysical data from the Indian

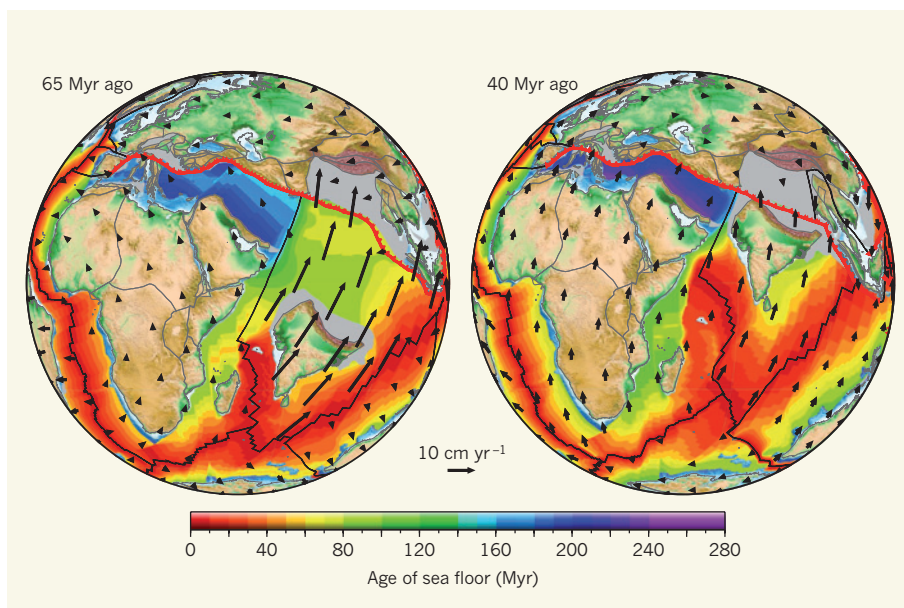


Figure 1 | The Indian and African plates 65 and 40 million years ago. The black arrows represent absolute plate motions, with the fast motion of India 65 Myr ago, and the slow motion of Africa, corresponding to a time shortly after the Réunion plume head is thought to have arrived at Earth's surface; the plume head is hypothesized¹ to have affected the plates' speeds. By 40 Myr ago, the influence of the plume had waned: the Indian plate had slowed down considerably and the African plate had resumed its former direction and speed. Rotated present-day topography is shown on the continents, and sea-floor age is depicted in the ocean basins. The estimated extent of now-destroyed continental margins is represented by the grey areas; mid-ocean ridges and transform faults are shown as black lines; subduction zones are shown as a red line, with teeth on the overriding plate. (Plate reconstructions generated using the GPlates software, data and rotations¹¹.)

Ocean, they note several phenomena associated with this event: the Indian plate sped up between 68 and 66 Myr ago and subsequently sustained absolute plate-motion speeds of 10–12 centimetres per year for about 15 Myr, with its peak speed reaching about 18 cm yr⁻¹. Between 52 and 45 Myr ago, India's motion slowed dramatically to less than 4 cm yr⁻¹. In addition, Africa's motion slowed down while India's sped up, only to return to its previous path and speed after India had decelerated. Figure 1 shows the state of play at 65 and 40 Myr ago, with respective fast and slower motion of the Indian plate.

India's superfast motion has been puzzling. One model ascribed it to loss of India's continental roots owing to mantle-plume activity when the supercontinent Gondwana originally broke apart⁵. However, such an erosion and

melting away of India's roots, thought to have occurred around 130 Myr ago, cannot be held responsible for the specific acceleration starting 67 Myr ago. Moreover, some geodynamic models suggest that the maximum velocity that tectonic plates typically reach is 8 cm yr⁻¹. The reason for this maximum speed is that plates are largely driven by the pull of slabs subducting in the mantle, especially upper-mantle slabs to which plates are attached^{6,7}. The characteristic sinking velocity for upper-mantle slabs is about 7 cm yr⁻¹, with a maximum of 8 cm yr⁻¹ for the oldest, thickest, least buoyant slabs⁶. This work has been confirmed by sophisticated numerical models for present-day plate motions⁸, which show a near-perfect agreement between observed plate motions and those predicted if plates are driven only by upper-mantle slabs.

Therefore, geodynamic models raise the question of how plates can achieve velocities in such great excess of 8 cm yr^{-1} . Cande and Stegman¹ use their detailed kinematic models to make a conceptual argument that, by adding a plume-push force, India can be sufficiently accelerated while Africa is slowed down. This dichotomy is explained by considering the location of the Réunion plume relative to the two plates. In India's case, the plume-push force would act in unison with slab pull north of India, whereas in Africa's case it would counteract slab pull, slowing the continent down. After exhaustion of the plume head and waning of its presumed pushing force, Africa would exhibit a speeding up synchronous with the slowdown of India.

Of course, this scenario is not without problems. One potential difficulty is that the Deccan traps — emplaced as a direct consequence of the partial melts supplied by the Réunion plume head — were generated over a comparatively short period of time, between 68 and 65 Myr ago⁴. On the other hand, the velocity anomaly of the Indian plate lasted for more than 15 Myr. Additional evidence for a relatively short-lived thermal anomaly under India comes from a combination of seismic images of the crust offshore of southwest India and computer simulations of crustal stretching. This leads to the conclusion that, by 63 Myr ago, relatively thin ocean crust was produced in the vicinity of the Réunion plume, suggesting a significant reduction in its thermal anomaly by that time⁵. One would imagine that, to satisfy Cande and Stegman's conceptual model¹, the plume head and its pushing force had to persist for a further 11 Myr, until 52 Myr ago when India slowed down substantially. After all, plates have no momentum — if a driving force is taken away, they should slow down relatively quickly.

Another potential problem comes from geodynamical modelling. Van Hinsbergen *et al.*¹⁰ have explored the magnitude of the plume-push force and concluded that although a plume-head effect could account for an acceleration of India, it could do so by only a few centimetres per year, certainly not the 10 cm yr^{-1} required to explain the Indian-plate velocities. Equivalently, their models indicate that a combination of reduced plume flux, reduced coupling with the plate and movement of the plate away from the plume can account for only a small portion of India's dramatic slowdown after 52 Myr ago.

Despite these problems, Cande and Stegman's work¹ stands out as an elegant demonstration of how plate kinematic data alone can be used to make inferences about the histories of coupling and synchronous motion of adjacent plates. Whether the plume-push force model for India's superfast motion will stand the test of time remains to be seen, but it offers an intriguing solution to a long-standing dilemma. ■

R. Dietmar Müller is in the School of Geosciences, University of Sydney, New South Wales 2006, Australia.
e-mail: dietmar.muller@sydney.edu.au

1. Cande, S. C. & Stegman, D. R. *Nature* **475**, 47–52 (2011).
2. Coffin, M. F. & Eldholm, O. *Rev. Geophys.* **32**, 1–36, doi:10.1029/93RG02508 (1994).
3. Torsvik, T. H., Burke, K., Steinberger, B., Webb, S. J. & Ashwal, L. D. *Nature* **466**, 352–355 (2010).
4. Chenet, A.-L., Quidelleur, X., Fluteau, F., Courtillot, V. & Bajpai, S. *Earth Planet. Sci. Lett.* **263**, 1–15 (2007).

5. Kumar, P. *et al. Nature* **449**, 894–897 (2007).
6. Goes, S., Capitanio, F. A. & Morra, G. *Nature* **451**, 981–984 (2008).
7. Schellart, W. P., Freeman, J., Stegman, D. R., Moresi, L. & May, D. *Nature* **446**, 308–311 (2007).
8. Stadler, G. *et al. Science* **329**, 1033–1038 (2010).
9. Armitage, J. J., Collier, J. S., Minshull, T. A. & Henstock, T. J. *Geochim. Geophys. Geosyst.* **12**, Q0AB07, doi:10.1029/2010gc003316 (2011).
10. van Hinsbergen, D. J. J., Steinberger, B., Doubrovine, P. V. & Gassmöller, R. *J. Geophys. Res.* **116**, B06101, doi:10.1029/2010JB008051 (2011).
11. www.gplates.org

STRUCTURAL BIOLOGY

A platform for copper pumps

Copper is vital to most cells, but too much is lethal. The structure of a protein that pumps copper ions out of the cytosol provides insight into both the pumping mechanism and how certain mutations in the protein cause disease. [SEE ARTICLE P.59](#)

NIGEL J. ROBINSON

Fifty years ago, the neurologist John Menkes described an inherited brain disease in children that was associated with brittle, white, kinky hair. A clue to the origin of the fatal condition came from sheep grazing on copper-deficient pasture in Australia: the animals' steely wool looked like the hair of children with Menkes disease. In both cases, the explanation was found to be inactive copper-dependent enzymes (cuproenzymes). More precisely, affected children have a faulty protein¹ known as the P1B-type ATPase ATP7A (often called the Menkes ATPase), an enzyme that pumps copper ions through membranes and has a vital role in distributing copper around the body.

On page 59 of this issue, Gourdon *et al.*² present the much-anticipated crystal structure of a P1B-type ATPase, and reveal a platform in the protein that creates a loading site for copper. A helix peculiar to P1B-type ATPases bends to form the platform, and the authors find that amino-acid residues at the bend of this helix are often mutated in the ATP7A protein of patients with Menkes disease (Fig. 1). More broadly, the structure provides information about the architecture of the P1B-type ATPase family of metal-transporter proteins, which are ubiquitous in bacteria, plants and animals.

Cells handle copper with care. Most cuproenzymes are confined to intracellular compartments or are expelled from cells to perform roles in extracellular fluids and matrices. Other proteins known as metallochaperones 'hand-deliver' copper to where it is needed, relinquishing their

load only to those proteins with which they specifically interact and which bind copper most tightly³. Proteins such as the Menkes ATPase pump copper from cells to prevent an excess of the metal from accumulating inside. They also pump copper from the cytosol into intracellular compartments where the metal is needed as a cofactor for cuproenzymes⁴, as well as into compartments that carry copper and cuproenzymes out of cells. When the Menkes ATPase fails, rather than being loaded into the bloodstream, copper becomes lodged inside cells lining the intestine. This, in turn, leads to a defective supply of copper cofactor for cuproenzymes⁵.

Gourdon and colleagues' structure² of LpCopA — a P1B-type ATPase from the bacterium *Legionella pneumophila* — is relevant to an eclectic range of sub-disciplines. For example, plant P1B-type ATPases deliver copper into the compartments of chloroplasts, where the metal ions conduct electrons in the process that converts light into chemical energy (the first stage of photosynthesis)⁶. Elevated levels of the Menkes ATPase have been linked to the progression of Alzheimer's disease, and a related protein, ATP7B, limits the efficacy of the anticancer drug cisplatin. P1B-type ATPases are also vital weapons in the immune system's battle against pathogenic bacteria⁷: when immune cells engulf invading microbes, they attack them with bactericidal copper pumped by Menkes ATPases; the bacteria use their own P1B-type ATPases in defence. Microbial P1B-type ATPases also counter the copper used as a fungicide in vineyards, and as a disinfectant in public water systems that can harbour Legionnaires' disease.

Because copper is bound and/or chaperoned

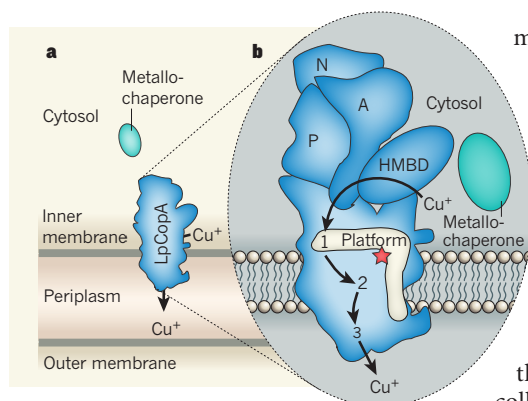


Figure 1 | A copper-exporting enzyme. **a**, Gram-negative bacteria use P1B-type ATPase enzymes (such as LpCopA in *L. pneumophila*, shown) to pump copper ions from the cytosol across their inner membrane and into the periplasm. In many organisms, metallochaperone molecules deliver the copper to the enzyme. **b**, Gourdon and colleagues' crystal structure² of LpCopA reveals a platform formed by a kink in an α -helix of the protein. The structure suggests that copper's pathway through the protein (arrows) involves three sets of binding sites (numbered 1 to 3), the first of which is on the platform. Copper is probably delivered to the platform site by a metallochaperone (as shown), or by a mobile heavy-metal-binding domain (HMBD) of LpCopA. The HMBD might regulate the activity of LpCopA by interacting with the protein's actuator domain (A) or nucleotide-binding domain (N), or, indirectly, with the phosphorylation domain (P). By analogy with the LpCopA structure, the authors deduce that certain mutations of the ATPase ATP7A that are associated with Menkes disease cluster around the kink of the platform-forming helix (red star).

in cells, there has been interest in understanding how P1B-type ATPases acquire copper from the cytosol. These proteins have soluble heavy-metal-binding domains (HMBDs, Fig. 1) that protrude into the cytosol and interact with copper chaperones electrostatically. At first, HMBDs were presumed to channel copper to the throat of the pump, but this model has been questioned⁸.

Gourdon and colleagues' structure² of LpCopA casts fresh light on this issue. The platform that they have discovered in the protein has positive charges around the putative entry site for copper, providing an ideal docking surface for a negatively charged copper donor. But is the donor a copper chaperone, the HMBD or both? This is a difficult question to answer for LpCopA, in part because the mobility of the HMBD prevented the authors from obtaining a fully resolved structure of this region. Nonetheless, the discovery of the platform in LpCopA should enable the identification of platform regions in other P1B-type ATPases for which the structures of isolated HMBDs and copper chaperones are already known, and this should, in turn, allow the question to be answered. Meanwhile, the

mobility of the HMBD suggests that it is well suited to regulating the activity of LpCopA — by, for example, forming interactions with the nearby actuator domain of the protein (Fig. 1b).

LpCopA spans the inner cell membrane of *L. pneumophila*, and pumps copper from the cytosol to the periplasm (the region between the inner and outer membranes; Fig. 1). The protein undergoes cycles of conformational changes that drive the passage of the metal. Gourdon and colleagues' structure² supplies crucial evidence for how this might occur. The authors observe that three amino-acid residues on the platform of LpCopA are sufficiently close together to simultaneously bind copper ions. They suggest that two of these residues pull apart from the third as the protein shifts into the next conformational state of the pumping cycle. At the same time, other copper-binding amino-acid residues deeper in the membrane approach the third original copper-binding residue. These combined conformational changes should attract copper ions from the platform to binding sites in the transmembrane region of LpCopA. In the final stage of the pumping cycle, the authors propose that the enzyme reverts to its initial state, whereupon a crucial copper-binding residue deep in the transmembrane region is 'tugged' by hydrogen bonds, allowing copper ions to be released to a binding site close to the outer face of the membrane. This site binds the metal only weakly, thus allowing the ion to escape into the periplasm.

P1B-type ATPases that selectively transport the biologically essential metals zinc and cobalt are known, as are others that transport the non-essential metals cadmium and lead. What enables these enzymes to select and pump metals other than copper? One probable answer is that the residues that form metal-binding sites in these proteins — those equivalent to the copper-binding residues in LpCopA — have been replaced with others through evolution, optimizing selective binding to the preferred metal. Indeed, some likely candidates for these alternative metal-binding residues have already been identified by comparison of the proteins' amino-acid sequences with those of copper-pumping enzymes. The charge on the metal-binding platform in these proteins might also repulse, rather than attract, copper chaperones. Alternatively, the HMBDs in the proteins might prevent copper chaperones either from reaching the platforms, or from regulating the activities, of these metal-ion pumps.

Many Gram-negative bacteria have a copper-efflux system that straddles both their inner and outer cell membranes, and exports copper from the cell by means of a ladder of copper-binding residues⁹. So why do these bacteria also need P1B-type proteins that

pump copper only part of the way out of the cell (that is, into the periplasm)? Indeed, why do many bacteria have several different kinds of P1B-type pumps for copper?

One view is that some P1B-type ATPases supply copper as a cofactor to cuproproteins in the periplasm. An alternative view is that some of these proteins pump copper ions from the periplasm into the cytosol — in which case, the conformational changes that generate a unidirectional drive in LpCopA would have to be modified to allow copper ions to flow in the opposite direction. A subset of pumps that have especially high affinities for copper but poor rates of outward transport has been identified, and members of this group are possible candidates for these alternative roles¹⁰. Gourdon and colleagues' structure should make it possible to identify parts of the copper-transport pathway that differ in this subset of P1B-type ATPases.

The ATPase ATP7B is found in liver cells, where it loads excess copper into bile ducts for excretion⁵. Mutations in ATP7B cause Wilson disease, symptoms of which include failure of copper excretion into bile and accumulation of the metal in the liver. The disease is managed by controlling copper levels in the body, for example, by administering the drug penicillamine, which binds to copper ions. Gourdon and colleagues' structure² of LpCopA should allow the biochemical consequences of different mutations associated with Wilson disease to be defined, possibly providing opportunities to develop personalized metal-management therapies. Furthermore, occipital horn syndrome and distal motor neuropathy are diseases that result from 'milder' mutations in the Menkes ATPase — the mutations cluster at the surface and extracellular face of the ATPase, unlike Menkes-disease mutations, which cluster near core transport functions of the copper pump². The authors' findings might also provide fresh avenues of investigation to improve treatments for these diseases. ■

Nigel J. Robinson is at the Biophysical Sciences Institute, Department of Chemistry, School of Biological and Biomedical Sciences, Durham University, Durham DH1 3LE, UK. e-mail: nigel.robinson@durham.ac.uk

1. Davies, K. *Nature* **361**, 98 (1993).
2. Gourdon, P. et al. *Nature* **475**, 59–64 (2011).
3. Banci, L. et al. *Nature* **465**, 645–648 (2010).
4. Setty, S. R. G. et al. *Nature* **454**, 1142–1146 (2008).
5. Kim, B.-E., Nevitt, T. & Thiele, D. J. *Nature Chem. Biol.* **4**, 176–185 (2008).
6. Abdel-Ghany, S. E., Müller-Moulé, P., Niyogi, K. K., Pilon, M. & Shikanai, T. *Plant Cell* **17**, 1233–1251 (2005).
7. White, C., Lee, J., Kambe, T., Fritzsche, K. & Petris, M. J. *J. Biol. Chem.* **284**, 33949–33956 (2009).
8. González-Guerrero, M. & Argüello, J. M. *Proc. Natl Acad. Sci. USA* **105**, 5992–5997 (2008).
9. Su, C.-C. et al. *Nature* **470**, 558–562 (2011).
10. González-Guerrero, M., Raimunda, D., Cheng, X. & Argüello, J. M. *Mol. Microbiol.* **78**, 1246–1258 (2010).

CANCER

When antioxidants are bad

Reactive oxygen species (ROS) get a bad press, as evidenced by the notable trend in the use of dietary and cosmetic antioxidants. New work suggests, however, that ROS might have a role in mitigating certain cancers. [SEE LETTER P.106](#)

RUSHIKA M. PERERA
& NABEEL BARDEESY

High levels of reactive oxygen species (ROS) have been considered to promote cancer^{1,2}. Yet the levels of Nrf2 — the transcription factor that mainly regulates physiological antioxidant pathways — is also increased in some cancers. Writing on page 106 of this issue, DeNicola *et al.*³ provide evidence in mice that several oncogenes actively induce transcription of *Nrf2*, promoting a ROS detoxification program that is required for tumour initiation.

Tumour cells can bypass normal responses to cellular stress by switching on programs driven by oncogenic mutations and changes in metabolism that allow sustained growth even in environments where nutrients and oxygen are scarce. This highly deregulated state is accompanied by the generation of potentially toxic by-products, including ROS. Paradoxically, however, whereas high ROS levels are harmful to normal cells, they have long been thought to aid tumour development in several ways: by inducing DNA damage and so accelerating the rate of cancer-causing mutations; by activating inflammatory pathways; and by stabilizing the hypoxia-inducible factor — a key regulator of energy metabolism⁴. These cancer-promoting actions of ROS underlie the notion that use of antioxidants will reduce cancer risk.

But there is another side to this coin: although many cancer cell types have increased levels of ROS, to avoid death even they must restrict their ROS 'footprint' below a given threshold. Indeed, tumours can use adaptive mechanisms to keep their ROS burden within a range permitting growth and survival⁵ — findings that call for a more nuanced view of the impact of ROS on tumorigenesis.

One such adaptive mechanism is genetic targeting of the KEAP1–NRF2 pathway in humans. Under normal conditions, the repressor protein KEAP1 promotes NRF2 degradation. In response

to extrinsic or intrinsic cellular stress, however, NRF2 is stabilized, entering the nucleus to activate cytoprotective genes⁶. Indeed, the KEAP1–NRF2 pathway regulates more than 100 protective genes, including components of an antioxidant system that can balance high ROS levels⁷. Notably, mutations that cause NRF2 stabilization, and so confer permanent protection against oxidative stress, have been detected in subsets of human lung, head-and-neck, and gall-bladder cancers⁷. These

data strongly suggest that an increase in basal antioxidant capacity can contribute to cancer development. So why are mutations in *KEAP1* and *NRF2* relatively rare in cancers?

DeNicola and colleagues' insightful study³ now shows that the Keap1–Nrf2 pathway in mouse cells can also be activated by means other than mutations in *Keap1* or *Nrf2*. They find that the *K-Ras* oncogene can regulate *Nrf2* transcription through a Mek–Erk–Jun signalling pathway, and cause a corresponding reduction in ROS in primary fibroblasts. This effect depends on levels of oncogene activation: expression of K-Ras at 'endogenous' levels, akin to those observed as tumour cells develop from normal cells, reduces ROS levels, whereas huge increases in K-Ras expression induce cellular stress and increase ROS levels. Importantly, the authors present evidence in mice that an antioxidant program is induced in the early stages of K-Ras-driven pancreatic-cancer development. Moreover, they report that transcriptional activation of Nrf2 could be part of the tumorigenic activity of other oncogenes such as *Braf* and *c-Myc*, and therefore contribute to the development of other malignancies such as lung cancer.

Thus, this study highlights two important points: a stress-response program is turned on early in tumour development; and oncogene activity is coupled with the transcriptional induction of Nrf2 (Fig. 1). It is therefore possible that a larger spectrum of cancers utilize this non-mutational pathway at early stages, to create a reducing environment that enables tumour initiation.

DeNicola *et al.* also provide genetic support for their cell-based observations in a mouse model, elegantly showing that Nrf2 deficiency impairs the development of K-Ras- and Braf-driven tumours. What's more, early-stage cancer cells lacking Nrf2 had high ROS levels and exhibited a senescence-like growth arrest; but if they were treated with antioxidants, tumour proliferation resumed.

Nrf2 controls numerous stress-response genes beyond those mitigating oxidative stress; these include regulators of NADPH regeneration, heat-shock proteins, drug-efflux pumps and various growth factors⁷. It is therefore important to determine whether any of these alternative Nrf2 targets cooperate with the antioxidant response to facilitate tumorigenesis. Nevertheless, DeNicola and co-workers' results suggest that Nrf2-mediated adaptation to ROS stress is crucial for early stages of tumour development, and if blocked may slow cancer progression.

It is noteworthy that in previous studies, Nrf2-deficient mice showed increased, rather than decreased, sensitivity to the development of carcinogen-induced

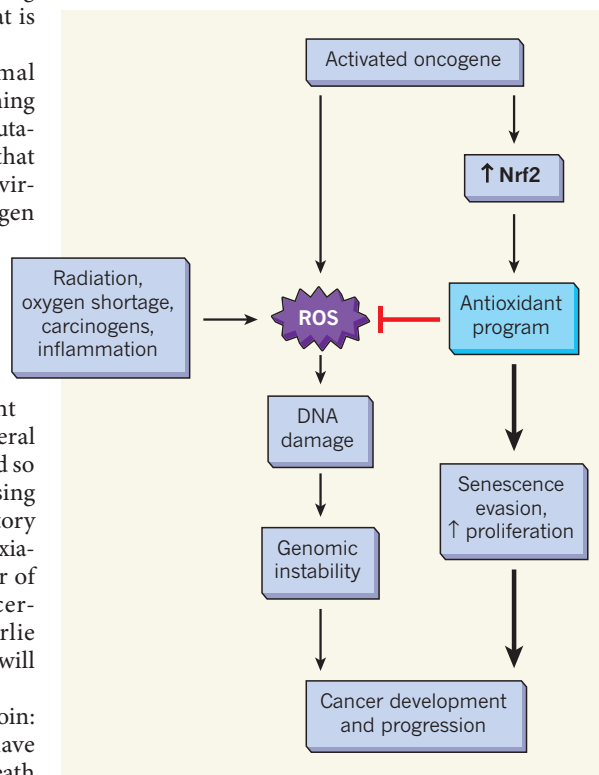


Figure 1 | The right balance of reactive oxygen species (ROS). Numerous intrinsic sources (such as certain activated oncogenes) and extrinsic factors (including radiation, oxygen shortage, carcinogens and inflammation) can give rise to ROS, ultimately leading to ROS-mediated genomic instability and cancer. In addition, cancer cells may simultaneously switch on protective mechanisms to balance the deleterious effects of chronically high ROS levels. DeNicola *et al.*³, for example, show that activation of the stress-response regulator NRF2 initiates a pre-emptive antioxidant program that can counter ROS as well as promote cell proliferation and survival. Tumour progression may therefore critically depend on the synergy between these two pathways.

cancers⁶, and that exposure to antioxidants delayed tumorigenesis in several animal models⁴. Moreover, a number of oncogenic lesions — for instance, loss of the tumour-suppressor protein p53 — are associated with increased ROS levels, which seems to drive tumorigenesis⁸. These observations highlight the context-dependent roles of ROS in cancer.

Along these lines, an open question is whether the KEAP1–NRF2 pathway and other active antioxidant processes also play a part during later stages of tumour development. In support of this idea, recent work⁹ indicates that growth of advanced pancreatic cancer requires ongoing autophagy — the process whereby damaged organelles and proteins are degraded — in part to suppress accumulation of cytotoxic ROS levels.

So what is the broader significance of DeNicola and colleagues' results³ for cancer prevention and therapy? In particular, do they imply that dietary antioxidants will increase the risk of cancer? The answer is probably no, because the authors found that antioxidant supplementation has a positive effect on tumorigenesis only in the setting of genetic deletion of Nrf2 — a context unlikely to be seen in human tumorigenesis.

Additional studies are required to define whether such supplements can either increase or decrease cancer risk depending on the presence of specific predisposing oncogenic lesions. As for the implications of these findings³ for cancer treatment, they further highlight the potential for success of clinical trials that aim to disable the protective antioxidant pathway¹⁰, thus making cancer cells particularly vulnerable to death by oxidative stress. So far, there are very few treatment options for patients with either pancreatic or lung cancers. Identifying an Achilles heel for these tumours may provide an opportunity to develop new treatments for these devastating diseases. ■

Rushika M. Perera and Nabeel Bardeesy are at the Massachusetts General Hospital, Cancer Center, and Department of Medicine, Harvard Medical School, Boston, Massachusetts 02114, USA.

e-mails: rperera@partners.org;
bardeesy.nabeel@mgh.harvard.edu

1. Vafa, O. *et al.* *Mol. Cell* **9**, 1031–1044 (2002).
2. Irani, K. *et al.* *Science* **275**, 1649–1652 (1997).
3. DeNicola, G. M. *et al.* *Nature* **475**, 106–109 (2011).
4. Gao, P. *et al.* *Cancer Cell* **12**, 230–238 (2007).
5. Gao, P. *et al.* *Nature* **458**, 762–765 (2009).
6. Kensler, T. W., Wakabayashi, N. & Biswal, S. *Annu. Rev. Pharmacol. Toxicol.* **47**, 89–116 (2007).
7. Hayes, J. D. & McMahon, M. *Trends Biochem. Sci.* **34**, 176–188 (2009).
8. Sablina, A. A. *et al.* *Nature Med.* **11**, 1306–1313 (2005).
9. Yang, S. *et al.* *Genes Dev.* **25**, 717–729 (2011).
10. Trachootham, D., Alexandre, J. & Huang, P. *Nature Rev. Drug Discov.* **8**, 579–591 (2009).

SATURN

Storm-clouds brooding on towering heights

Springtime on Saturn came in with a bang last December, when a massive storm erupted in its north temperate zone. This rare event has been observed in unprecedented detail from the ground and from space. SEE LETTERS P.71 & P.75

PETER READ

Saturn caught the attention of astronomers in early December 2010 when a small, prominent white spot appeared in its normally bland and hazy northern hemisphere. During the ensuing days, this spot grew in size to a diameter approaching that of Earth and developed a 'tail' of white clouds that spread eastwards until, two months later, it almost encircled the entire planet. Just five similar Great White Spot (GWS) events have been documented^{1,2} in the past 130 years. The events recurred at intervals of roughly Saturn's annual timescale of around 30 Earth years. What seems to be different about this occurrence are the unexpectedly early timing of its appearance (Saturn has only just begun its spring season in the northern hemisphere); its large size and powerful intensity; and the unprecedented observational resources available to measure and monitor its evolution.

In two papers published in this issue^{3,4} (pages 71 and 75), teams of observers, involving both ground-based astronomers (amateur

and professional) and those recording data from NASA's Cassini orbiter spacecraft, present some of the most detailed observations so far of such a dramatic event. These observations reveal with great clarity that the GWS comprises a massive complex of convective thunderstorms, upwelling heat energy and moisture from levels deep within Saturn's atmosphere, accompanied by huge and almost continuous lightning discharges.

Lightning discharges from convective storms were observed⁵ on Jupiter by direct imaging on the planet's nightside during the Voyager fly-bys in 1979. On Saturn, however, direct night-time imaging of lightning is not usually possible⁶ because the lightning flashes are too faint against the sunlight scattered from the planet's ring system. But pulses of radio emission, from events known as Saturn electrostatic discharges (SEDs), have been detected from Saturn and attributed to thunderstorm activity and lightning — for example⁷, from the much smaller 'dragon storm' observed by Cassini in 2004.

From the moment astronomers observed the

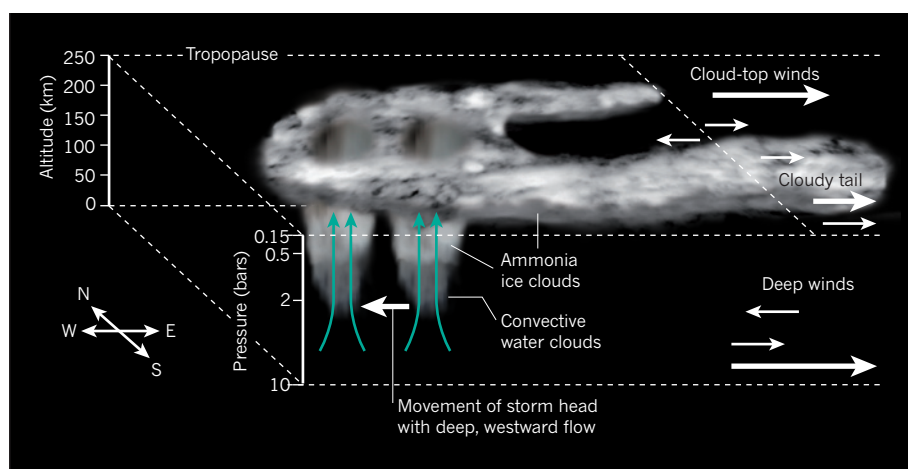


Figure 1 | Saturn's great storm of 2010–11. The studies of Sánchez-Lavega *et al.*³ and Fischer *et al.*⁴ indicate that the white clouds of this storm are produced by convective upwelling of heat, moisture and ammonia from deep water clouds (originating at pressure levels of about 10 bars) into the upper troposphere. As the upwelling material (green arrows) approaches the tropopause (where the temperature stops decreasing with height and begins increasing into the stratosphere) at pressures of 0.1–0.5 bars, the bright, white, ammonia ice clouds spread horizontally (like 'anvil' thunderstorm clouds on Earth) and get entrained into strong cloud-top eastward jets, producing an elongated 'tail', as visible in telescopic images of the storm. The cluster of convective clouds seems to move westward relative to the winds at the cloud tops, suggesting that westward zonal winds are stronger at deeper levels, where the convection is initiated.

first appearance of the white clouds at the onset of the present storm, Fischer *et al.*⁴ detected enormous bursts of SED activity using the Cassini Radio and Plasma Wave Science (RPWS) instrument. The intensity of each burst was around 10,000 times as strong as comparable discharges in terrestrial thunderstorms. But what was most remarkable about the present storm on Saturn was the incredible flash rate, peaking at 10 per second during the height of the storm. This indicates the presence of a multiple cluster of storm cells distributed over the approximately 5,000–8,000-kilometre-diameter convective region underlying the storm.

Latent heating associated with water condensation in convection has long been regarded⁸ as a key process for the upward transport of heat energy within the lowest portions of the atmosphere (the troposphere) of Saturn and Jupiter. Moisture-enhanced convection is so efficient at transporting heat that only a few intense, convecting columns of air may be needed to carry all of the upwelling heat energy from the deep interior (below the water clouds at pressures of around 10–20 bars, approximately 250 kilometres below the visible white-cloud tops) to the levels where this energy can be radiated to space (pressures of around 0.5 bars). As a result, deep, moisture-enhanced convection is highly intermittent on the two planets in both space and time.

Jupiter's moist-convection events seem to occur more or less randomly in time. This may reflect the small tilt (3.1°) of Jupiter's rotational axis, resulting in a relatively weak seasonal cycle. Saturn's axial tilt of 26.7° results in a much stronger seasonal modulation of the solar heating of its upper troposphere and stratosphere (the atmospheric layer just above the troposphere). However, direct sunlight is thought to penetrate only a short distance into the troposphere (perhaps down to pressures of just 1–2 bars), well short of the levels where water condensation and latent-heat release take place. So how does the seasonal cycle on Saturn control the onset of the deep convection events that produce a GWS?

The precise mechanisms at work here are still somewhat mysterious, although the observations and model simulations presented by Sánchez-Lavega *et al.*³ may provide some clues. Their observations indicate that the bright white clouds are produced when an upwelling convective plume of warm, moist air injects fresh ammonia ice particles into the upper troposphere (Fig. 1). From the growth in area of the clouds, the authors estimate upwelling wind speeds of several metres per second, although this may underestimate the speeds at deeper levels (which could exceed 100 m s⁻¹).

The location of the storm, close to the peak of a westward jet stream at 38°N and bordered north and south by strong eastward jets, is also significant and may account for the rapid eastward spread of its cloudy 'tail' in the sequence

of telescopic images (see Figs 1 and 2 of ref. 3). The westward drift of the storm, relative to the westward cloud-level jet within which the storm appeared, suggests an origin in the deep water clouds. The early stages of the ammonia injection and eastward spread of the upper cloudy tail seem quite clearly reproduced in Sánchez-Lavega and co-workers' numerical simulations, which at least confirm as plausible some of the assumptions (such as the very low static stability in the 'weather layer' between 10 and 0.5 bars) made in the model.

But many aspects of this phenomenon remain uncertain. Why did the storm appear at this particular latitude? It is intriguing that this latitude on Saturn coincides with the location of a chain of compact, enigmatic features (known as the 'string of pearls'^{1,9}) detected in the thermal infrared at depths of a few bars — could they be related to a hidden precursor disturbance that might have initiated the storm? What aspect of the seasonal cycle triggered the storm at this particular time? One possibility is perturbations to the thermal structure at the ammonia-cloud base (around 1–2 bars), caused by the strengthening of sunlight at the onset of spring. Such perturbations might have increased convective conditions there and allowed an upward extension of a pre-existing moist convective feature towards the top of the troposphere. Latent-heat release in an ammonia-rich, upwelling plume might have played a part in strengthening the convection¹⁰, although latent heating from ammonia condensation is usually negligible in Saturn's atmosphere.

It will be intriguing to see whether this event leads to significant, large-scale changes in Saturn's atmosphere as the storm evolves,

given that huge perturbations to Saturn's stratosphere are already detectable¹¹, and whether it might even prevent the onset of other stormy outbreaks near the summer solstice. All of these issues (and more) will clearly need further study using more sophisticated models and observations. But this latest event has already provided much new data to help inform and constrain these models.

Unlike the Hills of the Chankly Bore in the poem¹² by Edward Lear that inspired the title of this article, Saturn has no hills over which its storm clouds can brood, although, as Fischer and colleagues' observations⁴ show, these clouds host some pretty spectacular "fiery sparks". It will be a major challenge to the next generation of atmospheric models to predict when (and where) storm clouds will next appear to brood over Saturn's normally bland and hazy face. ■

Peter Read is in the Department of Physics, University of Oxford, Oxford OX1 3PU, UK. e-mail: p.read1@physics.ox.ac.uk

1. Del Genio, A. *et al.* in *Saturn from Cassini-Huygens* (eds Dougherty, M. K., Esposito, L. & Krimigis, S. M.) 113–159 (Cambridge Univ. Press, 2009).
2. Sánchez-Lavega, A. *et al.* *J. Geophys. Res.* **98**, 18857–18872 (1993).
3. Sánchez-Lavega, A. *et al.* *Nature* **475**, 71–74 (2011).
4. Fischer, G. *et al.* *Nature* **475**, 75–77 (2011).
5. Smith, B. A. *et al.* *Science* **204**, 951–972 (1979).
6. Dyudina, U. A. *et al.* *Geophys. Res. Lett.* **37**, L09205 (2010).
7. Dyudina, U. A. *et al.* *Icarus* **190**, 545–555 (2007).
8. Gierasch, P. J. *et al.* *Nature* **403**, 628–630 (2000).
9. <http://photojournal.jpl.nasa.gov/catalog/PIA01941>
10. Hueso, R. & Sánchez-Lavega, A. *Icarus* **172**, 255–271 (2004).
11. Fletcher, L. N. *et al.* *Science* **332**, 1413–1417 (2011).
12. Lear, E. 'The Dong with a Luminous Nose' in *Laughable Lyrics* (Robert John Bush, 1877).

METABOLISM

Alcohol, DNA and disease

Acetaldehyde, a reactive metabolite of ethanol, can damage DNA unless properly processed. A biochemical pathway involved in Fanconi anaemia seems to be essential for protection against such damage. SEE ARTICLE P.53

HANS JOENJE

Fanconi anaemia (FA) is a chromosome-breakage disease that causes developmental defects, sterility, bone-marrow failure and a highly elevated risk of cancer¹. Mutations in 15 genes are known to be associated with this disorder, and the encoded proteins are all thought to operate in the 'FA pathway'. This pathway preserves genome integrity by ensuring the correct repair of

certain types of DNA damage — in particular, damage caused by agents that generate DNA inter-strand crosslinks. But patients with FA have not necessarily been exposed to crosslinking agents, raising a long-standing question about the identity of the endogenous agents that generate FA-associated DNA damage. On page 53 of this issue, Langevin *et al.*² provide some answers. Somewhat alarmingly, they also shed light on how ethanol could lead to DNA damage.

A clue about potential endogenous DNA-damaging agents associated with FA came from work³ showing that cells carrying FA-associated mutations are exceptionally sensitive to the crosslinking agent formaldehyde, which — like most other aldehydes — is highly reactive. Aldehydes are omnipresent in nature and are generated during normal metabolism. There is considerable *in vitro* evidence that certain aldehydes can damage DNA, and can even cause crosslinking⁴. To prevent such damage, organisms have evolved a battery of aldehyde dehydrogenase enzymes that efficiently convert aldehydes into less noxious products⁵.

Langevin *et al.*² focused on the di-carbon aldehyde acetaldehyde, a degradation product of ethanol that, like several other alcohols, is itself produced during normal metabolism. They found that FA-mutant cell lines were highly sensitive to the addition of acetaldehyde, suggesting that endogenously generated acetaldehyde might be a driver of the features associated with FA.

The authors reasoned that if metabolically produced acetaldehyde is indeed a DNA-damaging agent normally counteracted by the FA pathway, then simultaneous absence of both the *Aldh2* gene (which encodes Aldh2, the main detoxifying enzyme of acetaldehyde) and the *Fancd2* gene — a key player in the FA pathway — should have a synergistic effect on the severity of the disease in mutant mice. (Note that, in contrast to humans with FA, mice with a defective FA pathway show hardly any overt symptoms, apart from reduced fertility and a mild susceptibility to cancer.)

When the authors attempted to generate double-mutant (*Aldh2*^{-/-} *Fancd2*^{-/-}) mice from *Aldh2*^{-/-} mothers, no such offspring were produced because the embryos were not viable. This suggests that when aldehyde detoxification is defective, the FA pathway becomes essential for embryonic development, and vice versa (Fig. 1). However, a more detailed genetic analysis revealed that *Aldh2*^{+/-} mothers, which can break down acetaldehyde, could give birth to live *Aldh2*^{-/-} *Fancd2*^{-/-} offspring, presumably because the developing embryos had access to maternal Aldh2 via the placenta. Nonetheless, the double-mutant mice were very unhealthy, often having developmental abnormalities and rapidly succumbing to acute leukaemia.

In addition to acetaldehyde, Aldh2 can break down a diverse range of other aldehydes, many of which are also generated naturally and can damage DNA. To confirm that

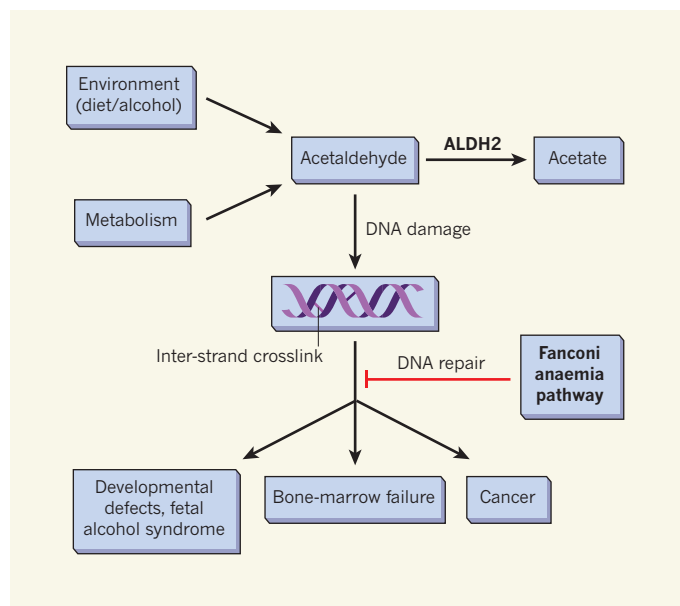


Figure 1 | Acetaldehyde metabolism and the Fanconi anaemia pathway of DNA repair. Acetaldehyde is generated naturally during metabolism, and its levels are further increased by external factors such as alcohol consumption. Normally, however, it is detoxified by aldehyde dehydrogenase enzymes such as ALDH2, which converts it into acetate. If this metabolic reaction is defective, acetaldehyde accumulation can damage DNA, causing inter-strand crosslinks. Langevin *et al.*² describe another layer of control in the form of DNA repair by the Fanconi anaemia pathway, which protects against effects of DNA damage such as developmental defects, bone-marrow failure and predisposition to cancer.

acetaldehyde accumulation alone is sufficient to cause disease in *Aldh2*^{-/-} *Fancd2*^{-/-} mice, Langevin *et al.* used ethanol, the immediate precursor of acetaldehyde. Exposure of pregnant mice to a single high dose of ethanol is known⁶ to cause fetal damage — it is an animal model for human fetal alcohol syndrome. The authors therefore tested whether a moderate dose of ethanol given to *Aldh2*^{+/-} pregnant mice affects the development of the double mutants. In fact, most of the resulting *Aldh2*^{-/-} *Fancd2*^{-/-} embryos died, and the few that did survive had severe developmental defects.

The researchers² then investigated the effect of ethanol exposure through drinking water on double-mutant mice born to mothers that had not been exposed to ethanol, and before they developed leukaemia. The effect was quite dramatic: the animals developed profound anaemia due to impaired bone-marrow function. This observation is particularly intriguing given that bone-marrow failure is a hallmark of FA in humans.

Langevin and colleagues' paper provides strong evidence that metabolically produced acetaldehyde is a potential driver of endogenous DNA damage, which is normally counteracted by acetaldehyde detoxification, in conjunction with DNA repair through the FA pathway (Fig. 1). It does not, however, address the question of whether acetaldehyde is the ultimate DNA-damaging agent. Although aldehydes can directly modify DNA chemically, it is possible that secondary factors are

also involved. These could include the depletion of substrates (NAD⁺, glutathione) needed for detoxification reactions, or the generation of reactive oxygen species⁷, which can also damage DNA if not detoxified⁸. Moreover, Langevin *et al.* focused on only one aldehyde. Much may be learned from similar investigations into how cells counteract the effects of other aldehydes, such as formaldehyde.

This work² provides a basis for understanding the pathogenesis of FA and may have consequences for treating individuals with the disease. For instance, patients could be advised to minimize their contact with sources of acetaldehyde and other aldehydes. Also, the symptoms of FA might be mitigated if patients' capacity for detoxifying aldehydes could be enhanced.

But perhaps most importantly, this study reinforces the great concern about the globally widespread consumption of alcohol. It underscores the idea that ethanol consumption is

genotoxic through the accumulation of acetaldehyde, potentially causing fetal damage (for instance, fetal alcohol syndrome), bone-marrow dysfunction and increased prevalence of cancers, particularly those of the oral cavity and oesophagus. Notably, an estimated 500 million people worldwide are deficient in ALDH2 activity and so are prone to acetaldehyde accumulation and subsequent genotoxicity⁹. Improved understanding of how ethanol can damage DNA may affect present attitudes to alcohol consumption. ■

Hans Joenje is in the Department of Clinical Genetics and the Cancer Center Amsterdam/VUmc Institute for Cancer and Immunology, VU University Medical Center, NL-1081 BT Amsterdam, the Netherlands.
e-mail: h.joenje@vumc.nl

1. www.fanconi.org
2. Langevin, F., Crossan, G. P., Rosado, I. V., Arends, M. J. & Patel, K. J. *Nature* **475**, 53–58 (2011).
3. Ridpath, J. R. *et al. Cancer Res.* **67**, 11117–11122 (2007).
4. Lorenti Garcia, C. *et al. Mutat. Res.* **662**, 3–9 (2009).
5. Marchitti, S. A., Brocker, C., Stagos, D. & Vasiliou, V. *Expert Opin. Drug Metab. Toxicol.* **4**, 697–720 (2008).
6. Sulik, K. K., Johnston, M. C. & Webb, M. A. *Science* **214**, 936–938 (1981).
7. Novitskiy, G., Traore, K., Wang, L., Trush, M. A. & Mezey, E. *Alcohol Clin. Exp. Res.* **30**, 1429–1435 (2006).
8. Joenje, H. *Mutat. Res.* **219**, 193–208 (1989).
9. Brooks, P. J., Enoch, M.-A., Goldman, D., Li, T.-K. & Yokoyama, A. *PLoS Med.* **6**, e50 (2009).

Indian and African plate motions driven by the push force of the Réunion plume head

Steven C. Cande¹ & Dave R. Stegman¹

Mantle plumes are thought to play an important part in the Earth's tectonics, yet it has been difficult to isolate the effect that plumes have on plate motions. Here we analyse the plate motions involved in two apparently disparate events—the unusually rapid motion of India between 67 and 52 million years ago and a contemporaneous, transitory slowing of Africa's motion—and show that the events are coupled, with the common element being the position of the Indian and African plates relative to the location of the Réunion plume head. The synchronicity of these events suggests that they were both driven by the force of the Réunion plume head. The recognition of this plume force has substantial tectonic implications: the speed-up and slowdown of India, the possible cessation of convergence between Africa and Eurasia in the Palaeocene epoch and the enigmatic bends of the fracture zones on the Southwest Indian Ridge can all be attributed to the Réunion plume.

A period of rapid motion of the Indian plate in the Late Cretaceous period and early Cenozoic era, starting around anomaly 30 (67 million years (Myr) ago) (ages are taken from ref. 1) and ending around anomaly 22 (49 Myr ago), was first recognized from magnetic anomalies on the Central Indian Ridge and Southeast Indian Ridge^{2,3} and later confirmed by palaeomagnetic data^{4–6}. The fast motion of the Indian plate during this time interval has been a major tectonic puzzle because the velocities of subducting plates are limited by their available plate driving forces—ridge-push and slab-pull—and because plates with continents on them tend to move more slowly than purely oceanic plates^{7,8}. An overlooked clue to this puzzle is that the onset of the superfast motion of India coincides with an unusual change in the motion of Africa: specifically, a large slowing of the rate of counterclockwise rotation of Africa about the Africa–Eurasia (Afr–Eur) Euler pole for that period (Fig. 1). A decrease in the Afr–Eur convergence rate at this time has previously been suggested from the summation of the plate circuit linking Africa to North America to Eurasia^{9,10}. However, this calculation has large uncertainties because of the complex tectonics in the North Atlantic Ocean^{11,12}.

The Réunion hotspot is an archetypal deep mantle plume with a plume head that arrived at the Earth's surface around 67 Myr ago, erupting the flood basalts that formed the Deccan traps, and leaving a hotspot track of islands and submarine ridges that formed as the Indian plate, and later the African plate, passed over the plume tail¹³ (Fig. 1). An enormous volume of basalts was erupted in a relatively short period of time around chron 29R (65 Myr ago)^{14,15}, although eruptions started as early as 67 Myr ago and continued until at least 62 Myr ago^{16,17}. After the initial massive eruption of basalts, volcanism began to wane and fell off considerably by 60–55 Myr ago^{18,19}. The plume head is related to regional tectonic disruptions, including the abandonment of the spreading centre in the Mascarene basin and the rifting of the Seychelles from India in the early Palaeocene epoch^{20–22}. However, the Réunion plume head has also been proposed to influence tectonics on a global scale, in particular the rapid northward motion of India during the Late Cretaceous and early Cenozoic^{23–25}.

Fast motion of the Indian plate

The motion of the Indian plate in the Late Cretaceous and early Cenozoic is best seen in the spreading-rate fluctuations of the India–Africa (Ind–Afr) and India–Antarctic (Ind–Ant) ridges

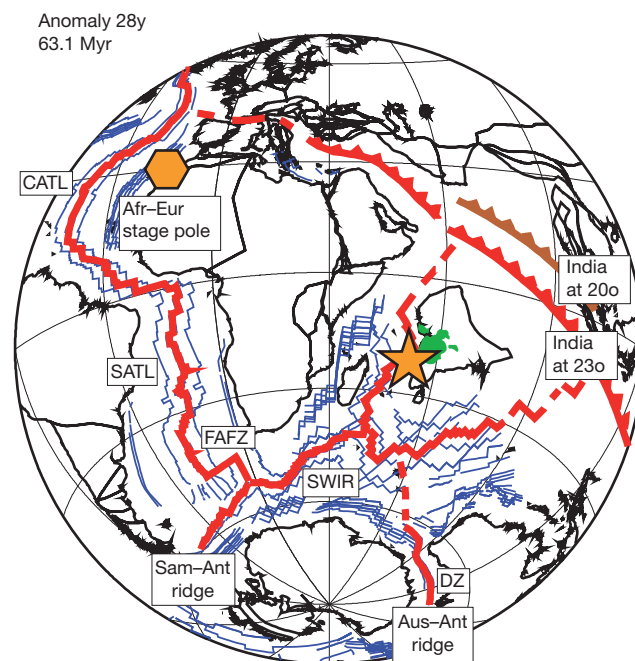


Figure 1 | Tectonic reconstruction of Indian Ocean at 63.1 Myr ago. The jagged red line (India at chron 23o) and the jagged brown line (India at chron 20o) show the positions of the trench between India and Eurasia at times of 'soft' and 'hard' collision, respectively, as determined by the reconstructed location of the northwest corner of India at those times. Yellow star, Réunion plume; green shading, Deccan flood basalts; yellow hexagon, the Afr–Eur stage pole; DZ, Diamantina zone.

¹Scripps Institution of Oceanography, La Jolla, California 92093-0220, USA.

(Fig. 2a) along two synthetic flowlines, as constrained by rotation parameters (see Fig. 2b for location). The spreading rates from chron 29 to chron 18 are derived from ref. 26 and the pre-chron 29 histories of the Ind–Afr and Ind–Ant ridges are constrained by refs 27 and 28, respectively. These flowlines reveal a short pulse (3 to 5 Myr) of superfast spreading (about 180 mm yr^{-1}) near the Cretaceous–Tertiary boundary, followed by a period of slower (although still fast) spreading (Ind–Ant = 130 mm yr^{-1} ; Ind–Afr = 110 mm yr^{-1}), which continued until roughly 52 Myr ago. The spreading rates then gradually decreased over a period of 7 Myr to much slower rates (Ind–Ant = 50 mm yr^{-1} ; Ind–Afr = 30 mm yr^{-1}).

A more detailed record of spreading rates around the Cretaceous–Tertiary boundary can be seen in two composite marine magnetic profiles, one across the fossil spreading centre in the Mascarene basin (Ind–Afr: anomalies 34 to 27) and the other across the same anomalies in conjugate corridors (that is, north and south flanks) of the Ind–Ant ridge. The spreading-rate histories based on these profiles (Fig. 2c, see Fig. 2b for locations) show that the pulse in fast spreading near the Cretaceous–Tertiary boundary was shorter and faster than suggested by the spreading-rate profiles on the basis of rotation parameters. Based on the timescale in ref. 1, the spreading rates reached a maximum within chron 29N on both ridges, at roughly 220 mm yr^{-1} on the Ind–Ant ridge and 200 mm yr^{-1} on the Ind–Afr ridge, and the period of superfast spreading lasted less than a million years.

We checked the widths of the polarity intervals between chrons 26 and 32 in ref. 1. Spreading-rate variations based on four representative magnetic profiles from the Pacific–Farallon (Pac–Far) ridge were computed at the same time intervals as used for the Indian Ocean

composite profiles. These four spreading-rate profiles are displayed above the magnetic polarity model in Fig. 2c. These profiles show that although the Pac–Far spreading rates are fairly constant for most of the time, there is a systematic slow/fast variation for chrons 29R and 29N, suggesting that the spacings of these chrons are slightly off. We found that adding 120,000 years to chron 29N and subtracting 120,000 years from 29R removed this spreading rate variation in the Pac–Far profiles (Fig. 2d), suggesting that this correction should be applied to the India Ocean profiles. With this correction we find that the pulse of superfast spreading occurred within chron 29R instead of 29N, was a little slower (Ind–Ant = 200 mm yr^{-1} instead of 220 mm yr^{-1}) and lasted longer, about 2.5 Myr.

The similarity between the motion of India and the eruptive history of the Deccan traps is remarkable. On the basis of the marine magnetic profiles (Fig. 2d), the pulse of superfast spreading started at around the beginning of chron 30 (67.8 Myr ago), roughly the same time as the first pulse of Deccan flood basalts, which is dated at 67 Myr ago¹⁷. The spreading rate rapidly increased and reached a maximum within chron 29R, the time of maximum basaltic eruptions. The spreading rate dropped off twice. The first decrease occurred around 63 Myr ago, by which time the main phase of Deccan volcanism had also ended. The spreading rate started to slow again around chron 23o (52 Myr ago) and continued to slow until chron 21y (45 Myr ago). This later slowing follows a drop-off in the eruptive volume of the Réunion plume^{18,19}.

The slowdown of the African plate

A transitory change in the motion of Africa in the Late Cretaceous and early Cenozoic is suggested by the large systematic variations in spreading rates on all of the adjacent ridges between 70 and 45 Myr ago. Figure 3a compares spreading-rate variations along flowlines in the Central Atlantic (CATL)²⁹, South Atlantic (SATL)²⁹ and Southwest Indian Ridge (SWIR)^{26,30} to the variations seen on the Ind–Afr ridge. Spreading rates in the SATL and SWIR both drop sharply between 70 and 65 Myr ago and continue to slow until they reach a minimum around 55 Myr ago, followed by an increase to roughly their pre-drop value at around 45 Myr ago. Spreading rates in the CATL also start to drop around 70 Myr ago, but then speed up and peak at 55 Myr ago.

These changes in spreading rates were accompanied by two large swings in the direction of spreading as recorded by fracture zones in all three oceans, one at the start of the period and one at the end^{29,31} (Fig. 3b). At the end of the period, the direction of spreading on all three ridges returned to what it was at the start. The simultaneous changes in motion on all three ridge systems indicates that it was the motion of Africa, the common plate, that was primarily changing. Africa's motion returned to its pre-slowness motion about the time that the fast motion of India ended, thus linking both of these events to the Réunion plume head.

The transitory behaviour of Africa can be quantified by an analysis of the systematic motion of the stage poles (Euler poles for a finite interval of time) on all three ridges between 70 and 45 Myr ago (Supplementary Table 1). As Fig. 4 shows, the stage poles for South America–Africa (Sam–Afr) and North America–Africa (Nam–Afr) motion²⁹, relative to a fixed African plate, follow very systematic paths between 70 and 45 Myr ago. Both sets of stage poles migrate along arcs that move directly away from the stage poles for chrons 34–32 (84–70 Myr ago) before returning to roughly the starting location. Intriguingly, if extended in the opposite direction these arcs pass very close to the Late Cretaceous–early Cenozoic stage pole of ref. 12 for Afr–Eur motion (chrons 30–21; yellow hexagon in Fig. 4). The stage pole path for Antarctica–Africa (Ant–Afr) motion^{26,30}, relative to a fixed Africa is not as well defined. However, the Ant–Afr stage pole for anomalies 29–25 also lies along an arc that passes through both the Ant–Afr stage pole for chrons 34–33 (84–73 Myr ago) (large green triangle in Fig. 4) and the Afr–Eur stage pole.

If the change in motion of Africa relative to the mantle is relatively simple, it will be expressed as a systematic variation in the differential

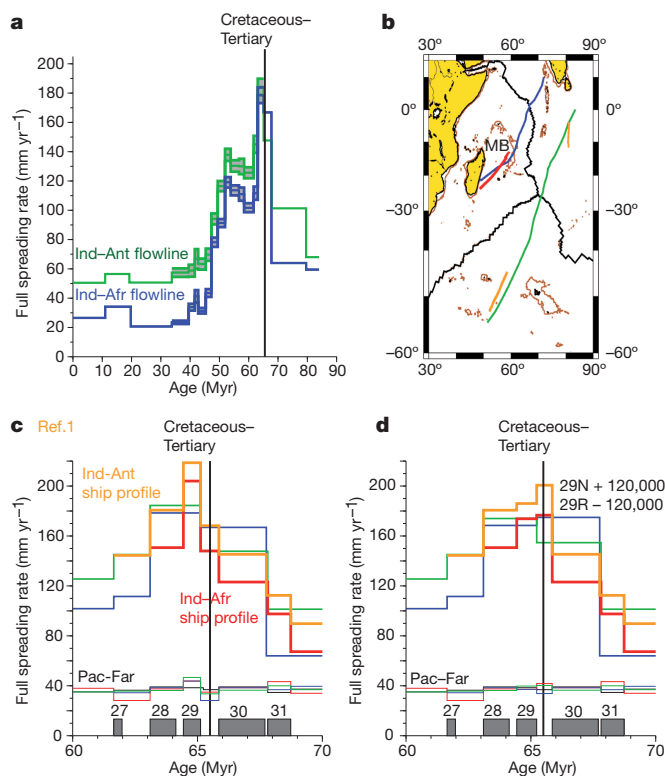


Figure 2 | Constraints on the motion of India. **a**, Spreading rates between Ind–Afr (blue) and Ind–Ant (green) ridges along synthetic flowlines. Green and blue bars between 35 and 65 Myr ago show 95% confidence intervals for spreading rates²⁶. **b**, Location of synthetic flowlines and shipboard profiles. MB, Mascarene basin. **c**, Comparison of spreading rates based on shipboard profiles (red and yellow) to synthetic flowlines (blue and green) for the period from 70 to 60 Myr ago. Profiles in the lower part of the figure show spreading rates for four magnetic lines from the Pac–Far ridge. The bar graph below shows chrons (numbers shown). **d**, Effect of adjusting lengths of chrons 29N and 29R to correct for systematic variation in Pac–Far spreading rates.

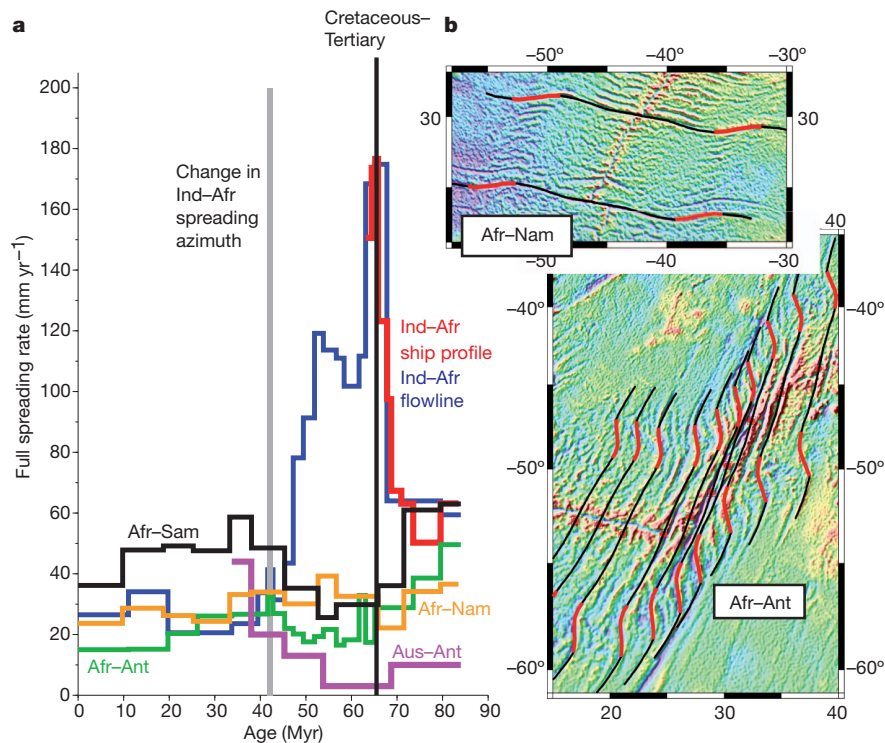


Figure 3 | Constraints on the motion of adjacent ridges. **a**, Spreading rates on the SATL (Afr–Sam), CATL (Afr–Nam), SWIR (Afr–Ant) and Australia–Antarctica (Aus–Ant) ridges since chron 34 (84 Myr ago). **b**, Synthetic flowlines

on the CATL (top panel) and SWIR (bottom panel) since chron 34, showing the large swings in spreading azimuths on both ridges between 70 and 45 Myr ago (red interval). Base maps show the Geosat gravity data of ref. 49.

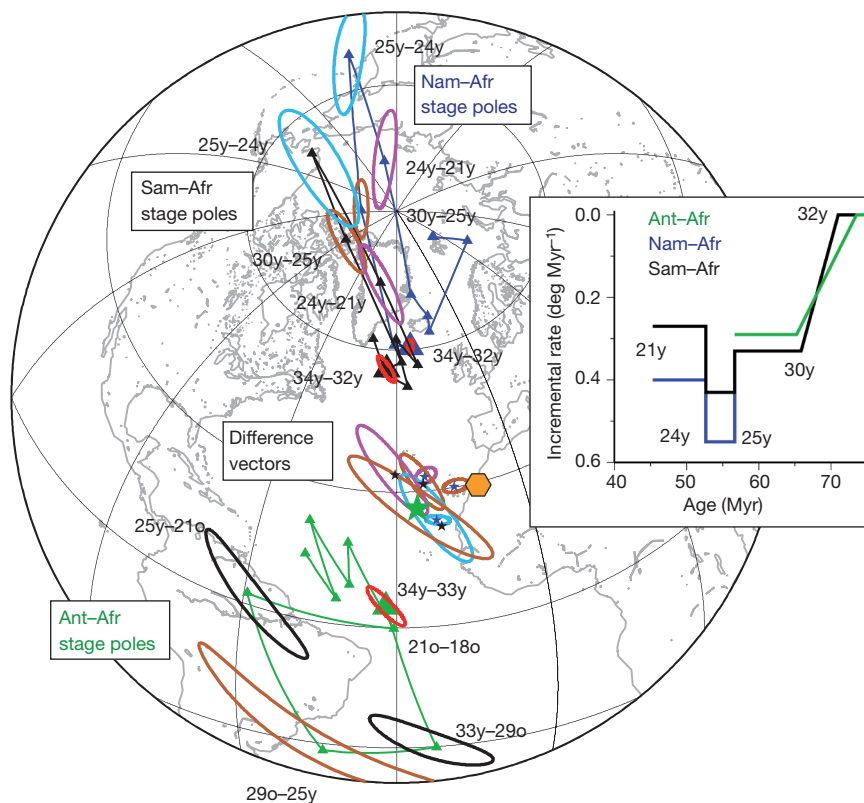


Figure 4 | Constraints on Africa's absolute motion. Difference vectors (stars) and stage poles (triangles) for the Nam–Afr (dark blue), Sam–Afr (black) and Ant–Afr (green) ridges. The difference vectors all cluster near the early Cenozoic stage pole for Afr–Eur motion (yellow hexagon). The oldest (starting) stage poles are indicated by the larger triangles. Ellipses are 95% confidence

limits. Error ellipses indicate 34y–32y (red), 30y–25y (brown), 25y–24y (light blue), 24y–21y (purple) and unused stage poles (black). For clarity, the error ellipses for younger rotations are omitted. All stage poles are plotted with respect to a fixed African plate. The inset shows the incremental slowing of Africa motion based on the difference vectors.

motion between stage poles over successive time intervals. When the displacements involved are relatively small, stage poles may be treated as velocity vectors. The differences between stage poles are also vectors (that is, difference vectors), but their length and pole location have a different interpretation (illustrated in Fig. 5). The analysis further assumes that during this time interval, changes in motions of the adjacent plates (South America, North America and Antarctica) relative to the mantle were small compared to the change in motion of Africa relative to the mantle.

We calculated difference vectors—the differences in the velocity vectors—for the three ridges for successive time intervals between chrons 30y and 21y (66 and 45 Myr ago) as the stage poles swung through their arcs, subtracting the younger velocity vectors from an initial, older velocity vector (the starting stage pole: chrons 34–32y for the CATL and SATL and chrons 34–33y for the SWIR) (Supplementary Table 2). The difference vectors (stars in Fig. 4) are located very close to the Euler pole for Afr–Eur convergence (yellow hexagon in Fig. 4) and remain essentially fixed during the early Cenozoic. This indicates that they provide estimates of both the change in motion of Africa with respect to the mantle and the change in the rate of Afr–Eur convergence.

The lengths of the difference vectors correspond to the rate of slowing (inset to Fig. 4) and show systematic variation during the swing of the arcs: their lengths increased as the stage pole moved to its maximum displacement along the arc, and then shortened as the stage pole retreated back to its starting position. Initially, between 66 and 55 Myr ago (chrons 30y and 25), the magnitude of slowing is roughly the same on all three ridges: 0.33° per Myr in the SATL and CATL and 0.29° per Myr on the SWIR. This is followed by the period of maximum slowdown, between 57 and 53 Myr ago (chrons 25–24), which corresponds to the maximal swing in the Nam–Afr stage poles (responsible for the short pulse of faster spreading on the CATL; yellow line in Fig. 3a). The peak magnitude of Nam–Afr and Sam–Afr is 0.55° per Myr and 0.45° per Myr, respectively, as observed by difference vectors for the NATL and SATL. The cause of this additional slowing is ambiguous because motion over this time interval on the SWIR is not resolved. If slowing occurred on the SWIR, then its presence on all three ridges would strongly suggest that Africa's slowing with respect to Eurasia bottomed at this time. Alternatively, if the slowing is only in association with the North America and South America plates, but not Antarctica, it might reflect some other forces acting on these plates that did not slow Africa's

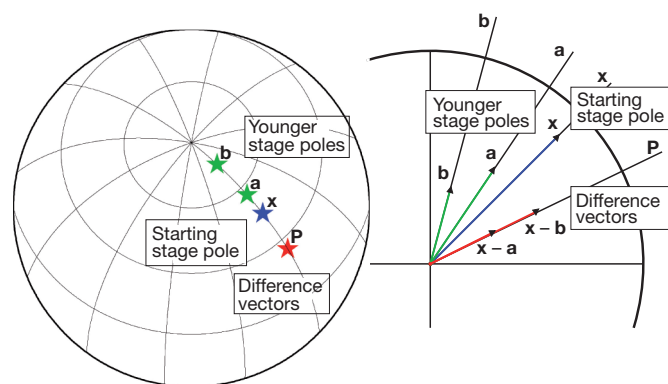


Figure 5 | Explanation of stage pole swing. Consider a sequence of stage poles for Africa's motion relative to an adjacent plate that starts at x , and migrates to a and then to b . In terms of Africa's motion, finding that the difference vectors, $x - a$ and $x - b$, stay in the same location P while their length is increasing means that the African plate is gradually slowing down about location P . The vectors $x - a$ and $x - b$ are being subtracted from Africa's motion. Although this incremental rotation is relative to the adjacent plate, if the change in motion of the adjacent plate relative to the mantle is small, then the incremental motion also gives the slowing of Africa relative to the mantle.

motion with respect to Eurasia and the mantle, such as additional braking due to compression across transform faults or perhaps the effect of the Iceland plume head, which reached the surface around chron 25 (ref. 21). Between chrons 24 and 21 the average incremental rate on the CATL and SATL reduces to about 0.34° per Myr. Between chrons 21 and 18 the stage pole swings end and plate motions return to their pre-swing configuration.

The difference vectors provide an indirect measurement of Afr–Eur convergence between 70 and 45 Myr ago if the incremental rate is subtracted from the total rate before or subsequent to Africa's slowdown. The only analysis of Afr–Eur motion in the Late Cretaceous and early Cenozoic with quantitative errors is in ref. 12, the authors of which summed Afr–Nam and Nam–Eur to get Afr–Eur motion. However, their oldest rotation was for anomaly 31, so we do not have a rate for a starting stage pole, say for anomalies 34 to 32, to which we can compare the slowing. We can, however, compare the difference vector rates to the average Afr–Eur rate in the post-swing period between chrons 21 and 13 (45 to 33 Myr ago) which is 0.35° per Myr (ref. 12). Using this as a background rate suggests that between chrons 30 and 25, Afr–Eur convergence essentially ceased.

If our model is correct, the stage poles for the absolute motion of Africa with respect to the mantle (Afr–Abs) should show the same systematic swing along an arc away from the Afr–Eur Euler stage pole as do the relative motion poles. Two models of absolute plate motion^{32,33} are displayed in Fig. 6a (blue and yellow lines), neither of which possesses the early Cenozoic stage pole swing in Africa's motion relative to the adjacent plates. We constructed an alternative absolute plate motion model for Africa for the last 84 Myr by taking three rotations from ref. 32 (chrons 5r, 18r and 34y) for the motion of Antarctica relative to the mantle (Ant–Abs), and summed them with the Ant–Afr relative motion rotations of refs 26 and 30 (Supplementary Table 3). Consequently, the resulting Afr–Abs stage poles (Supplementary Table 4) display the characteristic Africa stage pole swing between chrons 34–33 and chrons 29–25 (red line in Fig. 6a) and also yield a difference vector (Supplementary Table 5) near the Afr–Eur Euler pole (red star in Fig. 6a). We test this alternative model by comparing early Cenozoic hotspot tracks for the Walvis ridge and Great Meteor seamount with synthetic flowlines predicted by the model. Figure 6b shows that the curvature of the Walvis ridge between 25° S and 30° S (concave eastwards) is better fitted using this alternative model (red line) than using the previous model (blue line), as is the early Cenozoic path of the Great Meteor hotspot. There is little difference between the predicted hotspot tracks for Réunion (Supplementary Fig. 1).

Plume push driving force

We propose that a driving force originating from the Réunion plume head coupled the motions of India and Africa during the latest Cretaceous and early Cenozoic. The synchronous occurrence of a pulse of superfast spreading between 66 and 63 Myr ago with the maximum of Deccan flood basalt volcanism links the onset of India's fast motion to the Réunion plume head. This in itself does not establish a cause and effect relationship but is in sharp contrast to previous characterizations of the Réunion plume head being injected into an already fast-moving Indian plate (see, for example, ref. 34). The transitory slowing of the African plate with respect to the mantle also does not by itself argue for a plume force because there are many ways to slow convergence. However, the simultaneous occurrence of both of these unusual changes in plate motion, at the same time as the Deccan flood basalts, and in opposite directions from the location of the Deccan flood basalts, strongly suggests that their motions were influenced by a strong radial plate driving force emanating from the Réunion plume. Further, the approximate synchronicity of the slowdown of India between chrons 23 and 21 with the resumption of Africa's earlier motion strongly suggests that these later events are also coupled, and related to the waning strength of the plume.

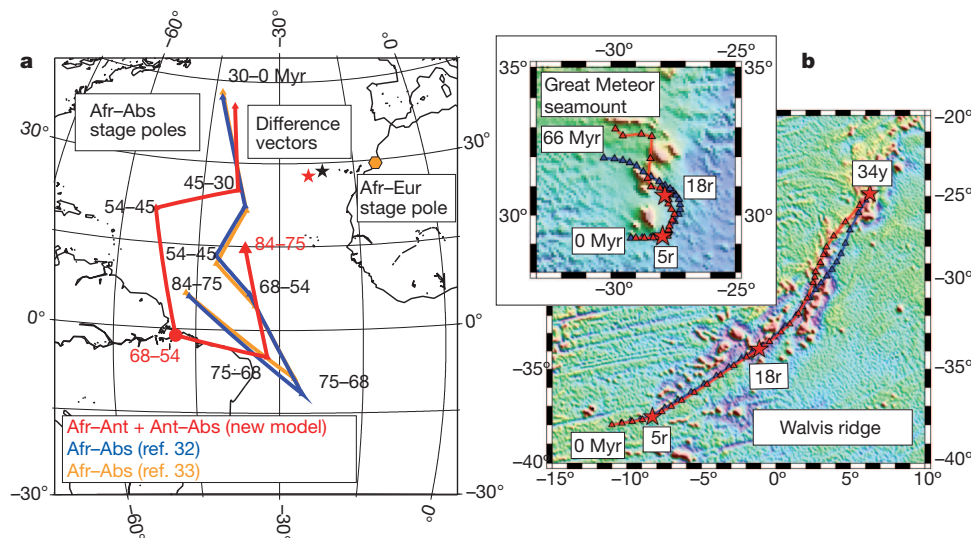


Figure 6 | Effect of the plume head force on absolute motion of Africa. **a**, Comparison of stage poles for three Africa hotspot models: Müller *et al.*³² (blue), Torsvik *et al.*³³ (yellow), and a new model (red) calculated by summing the relative motion of the Afr–Ant plates with three Ant–Abs rotations (chrons 5r, 18r and 34y) from ref. 32. The red star is the difference vector for the new model using the 68–54-Myr-ago stage pole and the 84–75-Myr-ago rotation as

a starting stage pole. The black star is the average of the six difference vectors for Nam–Afr and Sam–Afr. The yellow hexagon is the Afr–Eur stage pole. Stage poles are plotted with respect to a fixed African plate. **b**, Comparison of predicted hotspot trails for Tristan de Cunha (Walvis ridge) and the Great Meteor seamount, based on the Afr–Abs plate motion models: the new model (red) and the Müller *et al.* model³². Triangles mark 3-Myr increments.

The nature of the plume head force is poorly known. Previous work only considered buoyancy forces from hotspots (that is, plume conduits) in global models of plate driving forces (see, for example, ref. 35) and their net effect was found to be small. However, plume heads provide much greater buoyancy than their associated conduits and therefore, a greater driving force. Recent geodynamic models estimate that the forces generated by plume heads are unable to produce the 10 cm yr^{-1} speed-up observed for the Indian plate, and only account for $3\text{--}4 \text{ cm yr}^{-1}$ of the velocity increase³⁶. However, these models suggest that the waning strength of the plume leads to a gradual deceleration, which may be consistent with the slowdown we observe at 52–45 Myr ago.

Identifying the Réunion plume head as a major driving force for tectonic events between roughly 70 and 45 Myr ago has implications throughout the Indo-Atlantic region. If only the Eocene portion of Africa's transitory slowdown/speed-up is considered, the increase in Afr–Eur convergence rate since 55 Myr ago is missing its broader context. Our model suggests that this apparent speed-up simply represents a reversion to the pre-plume force balance rather than additional slab pull excited by a slab avalanche into the lower mantle³⁷. We also note that the Réunion plume is located 90° away from the Afr–Eur Euler pole (Fig. 1), in a position that allowed the plume push force to exert maximum torque on the plate, thus countering the force of the slab pull during the Palaeocene. Similarly, the major slowing of India during the Eocene epoch, which has generally been attributed to the initial collision of India and Eurasia^{38,39}, may also be reinterpreted. If the transitory slowdown of Africa reflects the time history of the driving force from the adjacent Réunion plume, India's gradual slowdown starting around chron 23o (52 Myr ago) may simply reflect the waning strength of the Réunion plume head. In this case, the actual collision might not have occurred until later, perhaps at chron 20o, the time of the 'hard' collision identified by ref. 39 (brown jagged line in Fig. 1; grey vertical line in Fig. 3a). Alternatively, if the collision occurred earlier, say around 55 Myr ago⁴⁰, the kinematic response of India to the Réunion plume head may have masked the kinematic evidence for the initial time of the India–Eurasia collision.

The slowing of Africa relative to the mantle placed many Central and South Atlantic transforms into compression. For example, the large Falkland–Agulhas fracture zone (FAFZ) in the South Atlantic,

with a 1,000-km left-stepping offset in the Late Cretaceous (Fig. 1), was put into compression at the start of Africa's slowdown; coincident with this event was a 700-km westward jump of the ridge and the abandonment of the spreading in the Agulhas basin⁴¹. The Sam–Ant ridge (Fig. 1) does a zigzag between chrons 30 and 24 similar to those on the SWIR and SATL^{30,42}, indicating that this event was caused by a transitory change in the relative motion between South America and Antarctica. The plume push force may have propagated across the Australia and Antarctic plates to cause the near-cessation of Australia–Antarctica (Aus–Ant) spreading (Fig. 1) between 70 and 54 Myr ago. The observed minimum in Aus–Ant spreading rates (3 mm yr^{-1} , Fig. 3a) is expressed in the enigmatic Diamantina Zone, which formed⁴³ between chrons 30 to 25, nearly synchronous with Africa's transitory slowdown.

The waning of the Réunion plume force between 52 and 45 Myr ago is also coincident with the age of the bend in the Hawaiian–Emperor chain⁴⁴, a feature which has been attributed to a combination of plate motion changes⁴⁵ and changes in mantle flow^{46,47}. It has been proposed that the cause of plate motion changes during this time is related to events in the western Pacific⁴⁸, but the ultimate cause for such changes is not known and we speculate that the waning strength of the Deccan plume head influenced plate motions as far as the Pacific.

METHODS SUMMARY

Spreading rates in Figs 2a and 3a were calculated from rotations in the cited publications along flowlines starting at 10° N , 72° E (Ind–Afr), 0° N , 81° E (Ind–Ant), 28° S , 34° W (Afr–Sam), 30° N , 45° W (Afr–Nam) and 40° S , 45° E (Afr–Ant). The 120,000-year correction to the width of chrons 29N and 29R in Fig. 2d was estimated by a trial-and-error fitting of the Pac–Far profiles. Difference vectors in Fig. 4 were calculated by treating the stage rotations as velocity vectors and calculating average rotation rates for each stage interval. The error ellipses for the difference vectors were calculated by scaling the corresponding covariance matrices for the stage rotations by the inverse of the square of the time interval in millions of years.

Received 6 January; accepted 9 May 2011.

1. Gradstein, F. M. *et al.* *A Geologic Time Scale 2004* (eds Gradstein, F. M., Ogg, J. G. & Smith, A. G.) (Cambridge University Press, 2004).
2. McKenzie, D. P. & Sclater, J. G. The evolution of the Indian Ocean since the Late Cretaceous. *Geophys. J. R. Astron. Soc.* **25**, 437–528 (1971).

3. Sclater, J. G. & Fisher, R. L. Evolution of the east central Indian Ocean, with emphasis on the tectonic setting of the Ninetyeast Ridge. *Geol. Soc. Am. Bull.* **85**, 683–702 (1974).
4. Besse, J. & Courtillot, V. Paleogeographic maps of the Indian Ocean bordering continents since the Upper Jurassic. *J. Geophys. Res.* **93**, 11791–11808 (1988).
5. Klootwijk, C. T., Gee, J. S., Peirce, J. W. & Smith, G. M. An early India–Asia contact: paleomagnetic constraints from Ninetyeast Ridge, ODP Leg 121. *Geology* **20**, 395–398 (1992).
6. Acton, G. D. in *Indian Subcontinent and Gondwana: A Palaeomagnetic and Rock Magnetic Perspective* (eds Radhakrishna, T. & Piper, J. D. A.) Vol. 44, 129–175 (Mem. Geol. Soc. India, 1999).
7. Jurdy, D. M. & Gordon, R. G. Global plate motions relative to the hotspots 64 to 56 Ma. *J. Geophys. Res.* **89**, 9927–9936 (1984).
8. Forsyth, D. & Uyeda, S. On the relative importance of the driving forces of plate motion. *Geophys. J. R. Astron. Soc.* **43**, 163–200 (1975).
9. Dercourt, J. *et al.* Geological evolution of the Tethys belt from the Atlantic to the Pamirs since the Lias. *Tectonophysics* **123**, 241–315 (1986).
10. Dewey, J. F., Helman, M. L., Turco, E., Hutton, D. H. W. & Knott, S. D. in *Alpine Tectonics* (eds Coward, M. P., Dietrich, D. & Park, R. G.) Vol. 45, 265–283 (Geol. Soc. Lond. Spec. Publ., 1989).
11. Srivastava, S. P. & Tapscott, C. R. in *The Western North Atlantic Region DNAG Vol. M* (eds Vogt, P. R. & Tucholke, B. E.) 379–404 (Geological Society of America, 1986).
12. McQuarrie, N., Stock, J. M., Verdel, C. & Wernicke, B. P. Cenozoic evolution of Neotethys and implications for the causes of plate motions. *Geophys. Res. Lett.* **30**, 2036, doi:10.1029/2003GL017992 (2003).
13. Duncan, R. A. & Richards, M. A. Hotspots, mantle plumes, flood basalts, and true polar wander. *Rev. Geophys.* **29**, 31–50 (1991).
14. Courtillot, V. E. *et al.* Deccan flood basalts at the Cretaceous/Tertiary boundary? *Earth Planet. Sci. Lett.* **80**, 361–374 (1986).
15. Vandamme, D., Courtillot, V., Besse, J. & Montigny, R. Paleomagnetism and age determinations of the Deccan traps (India): results of a Nagpur–Bombay traverse and review of earlier work. *Rev. Geophys. Space Phys.* **29**, 159–190 (1991).
16. Pande, K. Age and duration of the Deccan Traps, India: a review of radiometric and paleomagnetic constraints. *Proc. Indiana Acad. Sci.* **111**, 115–123 (2002).
17. Chenet, A. L., Quidelleur, X., Fluteau, F., Courtillot, V. & Bajpai, S. 40K–40Ar dating of the Main Deccan large igneous province: further evidence of KTB age and short duration. *Earth Planet. Sci. Lett.* **263**, 1–15 (2007).
18. White, R. Melt production rates in mantle plumes. *Phil. Trans. R. Soc. Lond. A* **342**, 137–153 (1993).
19. Tiwari, V. M., Grevemeyer, I., Singh, B. & Phipps Morgan, J. Variation of effective elastic thickness and melt production along the Deccan–Réunion hotspot track. *Earth Planet. Sci. Lett.* **264**, 9–21 (2007).
20. Norton, I. O. & Sclater, J. G. A model for the evolution of the Indian Ocean and the breakup of Gondwanaland. *J. Geophys. Res.* **84**, 6803–6830 (1979).
21. White, R. S. & McKenzie, D. P. Magmatism at rift zones: the generation of volcanic continental margins and flood basalts. *J. Geophys. Res.* **94**, 7685–7729 (1989).
22. Dymant, J. Structure et évolution de la lithosphère océanique dans l’océan Indien: apport des anomalies magnétiques. Thesis (Univ. Louis Pasteur, 1991).
23. Morgan, W. J. Deep mantle convective plumes and plate motions. *Am. Assoc. Pet. Geol. Bull.* **56**, 203–213 (1972).
24. Larson, R. L. Latest pulse of Earth: evidence for a mid-Cretaceous superplume. *Geology* **19**, 547–550 (1991).
25. Morgan, J. P., Morgan, W. J., Zhang, Y.-S. & Smith, W. H. F. Observational hints for a plume-fed, suboceanic asthenosphere and its role in mantle convection. *J. Geophys. Res.* **100**, 12753–12767 (1995).
26. Cande, S. C., Patriat, P. & Dymant, J. Motion between the Indian, Antarctica and African plates in the early Cenozoic. *Geophys. J. Int.* **183**, 127–149 (2010).
27. Molnar, P., Pardo-Casas, F. & Stock, J. The Cenozoic and Late Cretaceous evolution of the Indian Ocean: uncertainties in the reconstructed positions of the Indian, African and Antarctic plates. *Basin Res.* **1**, 23–40 (1988).
28. Royer, J. Y. & Sandwell, D. T. Evolution of the Eastern Indian Ocean since the Late Cretaceous: constraints from Geosat altimetry. *J. Geophys. Res.* **94**, 13755–13782 (1989).
29. Muller, R. D., Royer, J.-Y., Cande, S. C., Roest, W. R. & Maschenkov, S. in *Caribbean Basins* (ed. Mann, P.) Vol. 4, 33–59 (Elsevier Science, 1999).
30. Nankivell, A. P. Tectonic evolution of the Southern Ocean between Antarctica, South America and Africa over the last 84 Ma. PhD thesis (University of Oxford, 1997).
31. Royer, J.-Y., Patriat, P., Bergh, H. W. & Scotese, C. R. Evolution of the Southwest Indian Ridge from the Late Cretaceous (Anomaly 34) to the Middle Eocene (Anomaly 20). *Tectonophysics* **155**, 235–260 (1988).
32. Müller, R. D., Royer, J.-Y. & Lawver, L. A. Revised plate motions relative to the hotspots from combined Atlantic and Indian Ocean hotspot tracks. *Geology* **21**, 275–278 (1993).
33. Torsvik, T. H., Müller, R. D., Van der Voo, R., Steinberger, B. & Gaina, C. Global plate motion frames: toward a unified model. *Rev. Geophys.* **46**, RG3004 (2008).
34. Richards, M. A., Duncan, R. A. & Courtillot, V. E. Flood basalts and hotspot tracks: plume heads and tails. *Science* **246**, 103–107 (1989).
35. Harper, J. F. Plate dynamics: Caribbean map corrections and hotspot push. *Geophys. J. Int.* **100**, 423–431 (1990).
36. van Hinsbergen, D., Steinberger, B., Doubrovine, P. & Gassmoller, R. Acceleration and deceleration of India–Asia convergence since the Cretaceous: roles of mantle plumes and continental collision. *J. Geophys. Res.* doi:10.1029/2010JB008051 (in the press).
37. Capitanio, F. A., Faccenna, C. & Funicello, R. Opening of Sirte Basin: result of slab avalanche? *Earth Planet. Sci. Lett.* **285**, 210–216 (2009).
38. Molnar, P. & Tapponnier, P. Cenozoic tectonics of Asia: effects of a continental collision. *Science* **189**, 419–426 (1975).
39. Patriat, P. & Achache, J. India–Eurasia collision chronology has implications for shortening and driving mechanism of plates. *Nature* **311**, 615–621 (1984).
40. Garzanti, E., Baud, A. & Mascle, G. Sedimentary record of the northward flight of India and its collision with Eurasia (Ladakh Himalaya, India). *Geodin. Acta* **1**, 297–312 (1987).
41. Barker, P. F. The history of ridge-crest offset at the Falkland–Agulhas fracture zone from a small-circle geophysical profile. *Geophys. J. R. Astron. Soc.* **59**, 131–145 (1979).
42. Livermore, R. A. & Woollett, R. W. Seafloor spreading in the Weddell Sea and Southwest Atlantic since the Late Cretaceous. *Earth Planet. Sci. Lett.* **117**, 475–495 (1993).
43. Tikku, A. A. & Cande, S. C. The oldest magnetic anomalies in the Australian–Antarctic Basin: are they isochrons? *J. Geophys. Res.* **104**, 661–677 (1999).
44. Sharp, W. D. & Clague, D. A. 50-Ma initiation of Hawaiian–Emperor bend records major change in Pacific plate motion. *Science* **313**, 1281–1284 (2006).
45. Raymond, C. A., Stock, J. M. & Cande, S. C. In *History and Dynamics of Global Plate Motions* (eds Richards, M. A., Gordon, R. G. & van der Hilst, R. D.) *AGU Monogr.* **121**, 359–376 (2000).
46. Steinberger, B., Sutherland, R. & O’Connell, R. J. Prediction of Emperor–Hawaii seamount locations from a revised model of global plate motion and mantle flow. *Nature* **430**, 167–173 (2004).
47. Tarduno, J., Bunge, H.-P., Sleep, N. & Hansen, U. The bent Hawaiian–Emperor hotspot track: inheriting the mantle wind. *Science* **324**, 50–53 (2009).
48. Whittaker, J. M. *et al.* Major Australian–Antarctic plate reorganization at Hawaiian–Emperor bend time. *Science* **318**, 83–86 (2007).
49. Sandwell, D. T. & Smith, W. H. F. Marine gravity anomaly from Geosat and ERS-1 satellite altimetry. *J. Geophys. Res.* **102**, 10039–10054 (1997).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Stock for discussions. R. Gordon and D. Müller made comments on the manuscript. Funding was provided by NSF grant ANT-0944345 (to S.C.C.).

Author Contributions Both authors contributed equally to the ideas and design of the research. S.C.C. developed the new methodology and performed kinematic analysis. Both authors contributed to writing the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.C.C. (scande@ucsd.edu).

Fancd2 counteracts the toxic effects of naturally produced aldehydes in mice

Frédéric Langevin¹, Gerry P. Crossan¹, Ivan V. Rosado¹, Mark J. Arends² & Ketan J. Patel^{1,3}

Reactive aldehydes are common carcinogens. They are also by-products of several metabolic pathways and, without enzymatic catabolism, may accumulate and cause DNA damage. Ethanol, which is metabolised to acetaldehyde, is both carcinogenic and teratogenic in humans. Here we find that the Fanconi anaemia DNA repair pathway counteracts acetaldehyde-induced genotoxicity in mice. Our results show that the acetaldehyde-catabolising enzyme *Aldh2* is essential for the development of *Fancd2*^{-/-} embryos. Nevertheless, acetaldehyde-catabolism-competent mothers (*Aldh2*^{+/-}) can support the development of double-mutant (*Aldh2*^{-/-}*Fancd2*^{-/-}) mice. However, these embryos are unusually sensitive to ethanol exposure *in utero*, and ethanol consumption by postnatal double-deficient mice rapidly precipitates bone marrow failure. Lastly, *Aldh2*^{-/-}*Fancd2*^{-/-} mice spontaneously develop acute leukaemia. Acetaldehyde-mediated DNA damage may critically contribute to the genesis of fetal alcohol syndrome in fetuses, as well as to abnormal development, haematopoietic failure and cancer predisposition in Fanconi anaemia patients.

Multicellular organisms are constantly exposed to common environmental toxins that cause genome damage. In addition, both aerobic respiration and metabolism generate many reactive chemical species that are capable of attacking DNA¹. To counteract such endogenous threats to genome stability, cells catabolise these reactive molecules while simultaneously mobilizing DNA repair. The failure to respond to this threat leads to the catastrophic phenotypes caused by the genetic inactivation of certain DNA repair pathways. In Fanconi anaemia, biallelic mutations in any one of fourteen genes cause developmental defects, as well as progressive bone marrow failure and widespread cancer predisposition². In this specific clinical instance we have little or no insight into the source of endogenous DNA damage, and we do not yet understand how this eventually contributes to the development of this phenotype.

In searching for a source of endogenous DNA damage, we focused on simple aldehydes. These molecules are ubiquitous organic compounds—common constituents of food sources—and are also produced within organisms as by-products of cellular metabolism³. Aldehydes are very reactive molecules and are therefore likely to constitute a serious threat to cellular integrity. In fact, these molecules can react avidly with DNA *in vitro*, causing a range of DNA modifications^{4–6}. Whereas much is known about how aldehydes are detoxified in cells by a class of enzymes known as the aldehyde dehydrogenases^{7,8}, little is known about whether or not aldehydes damage DNA *in vivo* and, if so, how such DNA damage might be repaired. Here we present evidence indicating that the Fanconi anaemia DNA repair pathway has a crucial role in counteracting acetaldehyde-induced genotoxicity in mice.

Fanconi B cells are sensitive to acetaldehyde

A recent study tested chicken DT40 DNA repair mutant cell lines for hypersensitivity to formaldehyde⁹. The authors concluded that homologous recombination, translesion synthesis, and the Fanconi anaemia gene *FANCD2* protected against formaldehyde-induced clonogenic cell killing. Numerous studies have reported that cells exposed to acetaldehyde accumulate DNA damage^{10,11}. Moreover, ethanol exposure (a

dietary source of acetaldehyde) in mice and humans results in DNA base damage^{10–13}. We compared the sensitivity of a chicken B-cell line, DT40, carrying a disruption of the Fanconi anaemia gene *FANCL* (Δ *FANCL*) to acetaldehyde, and found that this mutant strain is markedly sensitive to exogenous acetaldehyde (Fig. 1a). Four additional Fanconi anaemia mutants that operate both upstream (Δ *FANCB* and Δ *FANCC*; Fig. 1b) and downstream in the pathway (Δ *FANCF* and Δ *RAD51C*; Fig. 1c)^{9,14,15} were also sensitive to this aldehyde. To determine further the genetic requirements for protection against acetaldehyde genotoxicity, we exposed various DNA repair mutants representing distinct DNA repair pathways to acetaldehyde. As shown in Fig. 1d and e, mutants in homologous recombination (Δ *XRCC2*), translesion synthesis (Δ *REV1*), nucleotide excision repair (Δ *XPA*), non-homologous end joining (Δ *KU70*; also known as *XRCC6*) and mismatch repair (Δ *EXO1*) are not particularly sensitive to acetaldehyde.

Developmental role for *Aldh2* in *Fancd2*^{-/-} mice

Although Fanconi anaemia pathway-deficient cell lines are hypersensitive to exogenous acetaldehyde, this does not address whether aldehyde toxicity occurs within a physiological context. This is significant as acetaldehyde is produced as a by-product of natural metabolism (Fig. 1f). Such intracellular pools of aldehydes are eliminated through the action of many enzymes, including numerous aldehyde dehydrogenases^{16,17}, aldehyde oxidase¹⁸ and cytochrome p450 (Fig. 1f)¹⁹. Although organisms use various means to catabolise aldehydes, we focused on the mitochondrial enzyme aldehyde dehydrogenase 2 because of its well-defined role in acetaldehyde catabolism in humans⁷. Surprisingly, disrupting *ALDH2* in Δ *FANCC* DT40 cells (Supplementary Fig. 1a) did not potentiate sensitivity to exogenous acetaldehyde (Supplementary Fig. 1b). We surmised that additional enzymes may compensate for *ALDH2* inactivation in these cells. However, the situation might be different if the same approach were tested in the context of a whole organism. Therefore, we set out to generate mice that harbour disruptions of the key Fanconi anaemia gene *Fancd2* in combination with *Aldh2*. *Fancd2* mice were chosen because they have the strongest phenotype among mice carrying

¹MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 0QH, UK. ²University of Cambridge, Department of Pathology, Addenbrooke's Hospital, Cambridge CB2 2QQ, UK. ³University of Cambridge, Department of Medicine, Level 5, Addenbrooke's Hospital, Cambridge CB2 0QQ, UK.

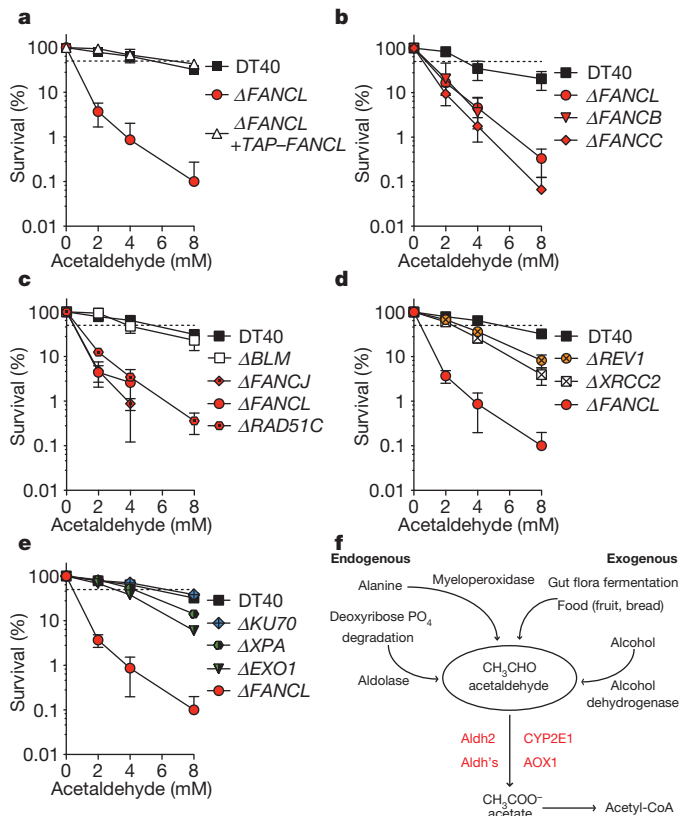


Figure 1 | Chicken Fanconi anaemia pathway knockout B cells are hypersensitive to acetaldehyde. **a–f**, Clonogenic survival of DT40 DNA repair mutants exposed to acetaldehyde. **a**, Fanconi anaemia: Δ FANCL and Δ FANCL complemented back with TAP-FANCL complementary DNA. **b**, Fanconi anaemia core complex components: Δ FANCB and Δ FANCC. **c**, Downstream Fanconi anaemia genes: Δ FANCL and Δ RAD51C. **d**, Homologous recombination repair: Δ XRCC2; and translesion synthesis: Δ REV1. **e**, Non-homologous end joining: Δ KU70; nucleotide excision repair: Δ XPA; and mismatch repair: Δ EXO1. In **a–e**, each point represents the mean of three independent experiments, each carried out in duplicate; error bars represent s.e.m. **f**, Scheme outlining the potential sources contributing to the endogenous pool of acetaldehyde^{42,43} and its catabolism.

mutations in Fanconi anaemia genes²⁰, with the exception of *Slx4* mice²¹. Furthermore, acetaldehyde exposure in cell lines induces the monoubiquitination of this key Fanconi anaemia protein²² (Supplementary Fig. 1c) and mouse *Fancd2*^{−/−} cells are sensitive to exogenous acetaldehyde (Supplementary Fig. 2).

First, we created *Aldh2*^{−/−} mice using embryonic stem cells harbouring a targeted mutation in the *Aldh2* locus obtained from the EUCOMM consortium (*Aldh2*^{tm1a(EUCOMM)Wtsi}; Supplementary Fig. 3a, b). Germline transmission of the knockout allele and subsequent crosses generated *Aldh2*^{−/−} mice in a C57BL/6J background. Homozygous mice were born at the expected Mendelian ratios and show no overt phenotype (Supplementary Fig. 3c), consistent with previously published *Aldh2* knockout mice²³. Second, we crossed *Aldh2*^{−/−} with *Fancd2*^{+/−} mice made in the 129S4 background²⁴. To obtain *Aldh2*^{−/−}*Fancd2*^{+/−} mice, we set up four types of crosses where two allelic configurations of the disrupted *Aldh2* locus (*Aldh2*^{+/−} or *Aldh2*^{−/−}) were combined with *Fancd2* heterozygosity (*Fancd2*^{+/−}) in both male and female mice (Fig. 2a). This was done because *Fancd2*^{−/−} mice are sterile. The observed genotypic frequencies presented in Fig. 2a show that when the maternal genotype was *Aldh2*^{−/−}, double-mutant offspring were not produced (associated with a reduction in mean litter size; Supplementary Fig. 4a). This is in contrast with the expected Mendelian segregation of 12.5% and 25% when the paternal genotypes are *Aldh2*^{+/−} and *Aldh2*^{−/−}, respectively (Fig. 2a

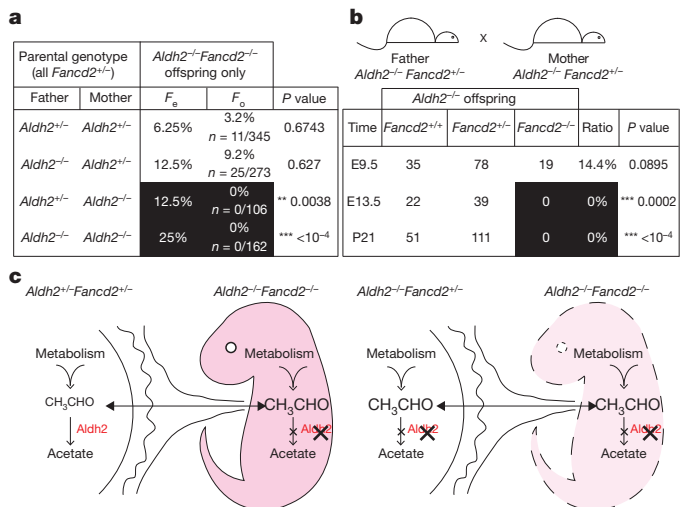


Figure 2 | A single maternal allele of *Aldh2* is essential for the development of *Aldh2*^{−/−}*Fancd2*^{−/−} mice. **a**, Crosses consisting of various allelic combinations at the *Aldh2* locus (always *Fancd2*^{+/−}) were set up to obtain *Aldh2*^{−/−}*Fancd2*^{−/−} pups. *F_e* and *F_o* are expected and observed Mendelian frequencies, respectively, of *Aldh2*^{−/−}*Fancd2*^{−/−} pups. Shaded regions highlight the cross for which the frequency is 0% when it is expected to be 12.5% or 25%. χ^2 test, confidence interval 5%. **b**, Observed frequencies of *Fancd2* genotypes in *Aldh2*^{−/−} embryos at E9.5, E13.5 and 21 days postpartum (P21) for the intercross outlined above the table. Each column represents the number of embryos obtained for each genotype. The last column defines the percentage of *Aldh2*^{−/−}*Fancd2*^{−/−} mice. Shaded region highlights cross for which the frequency is 0% when it is expected to be 25%. χ^2 test, confidence interval 5%. **c**, Schematic interpretation of how maternal or fetal *Aldh2* may contribute to the development of *Fancd2*^{−/−} mice. Light pink shading and dashed line indicate that the embryo dies.

and Supplementary Fig. 4b). However, when the females were *Aldh2*^{+/−}, then *Aldh2*^{−/−}*Fancd2*^{−/−} pups were born. Timed *Aldh2*^{−/−}*Fancd2*^{+/−} male mice \times *Aldh2*^{−/−}*Fancd2*^{+/−} female mice crosses were set up to determine when *Aldh2*^{−/−}*Fancd2*^{−/−} embryos died. Dams from such crosses were killed on embryonic days (E)9.5 and E13.5 and embryos were genotyped. At E9.5, *Aldh2*^{−/−}*Fancd2*^{−/−} embryos were observed at a frequency of 14.4% (Fig. 2b), but at E13.5 they had all died. This suggests that fetal attrition occurs between E9.5 and E13.5. These results indicate that the maternal *Aldh2* genotype profoundly influences the development of double-mutant embryos *in utero*. Figure 2c and Supplementary Fig. 4c illustrate the various situations whereby survival of *Fancd2*^{−/−} embryos is and is not supported; and the extent to which these results are dependent on maternal and fetal *Aldh2* genotypes. We conclude that, in the absence of the Fanconi anaemia pathway, either the mother or the fetus requires *Aldh2* to catabolise acetaldehyde to enable fetal development. Acetaldehyde is a volatile small molecule, thereby allowing it to passively diffuse across the placental membranes. This could then enable the mother to break down acetaldehyde, hence compensating for *Aldh2* deficiency in the fetus. Thus, a single *Aldh2* allele is sufficient to allow the development of *Fancd2*^{−/−} embryos.

Ethanol teratogenicity in *Aldh2*^{−/−}*Fancd2*^{+/−} mice

Next, we wanted to establish whether or not acetaldehyde toxicity was sufficient to cause fetal attrition. Previous studies have shown that C57BL/6J mice were susceptible to the teratogenic effects of ethanol, providing a mouse model for human fetal alcohol syndrome²⁵. This teratogenicity is largely due to acetaldehyde, a by-product of ethanol catabolism^{26,27}. Because we knew that an *Aldh2*^{−/−}*Fancd2*^{+/−} male \times *Aldh2*^{+/−}*Fancd2*^{+/−} female cross generated the highest frequency of viable double-mutant pups (\sim 9%; Fig. 2a, c), this cross was used to assess the impact of *in utero* exposure to acetaldehyde derived from

ethanol catabolism. It is noteworthy that the double-mutant pups born from this cross often carry subtle defects such as kinked tails and eye defects (Supplementary Fig. 5a–c). Figure 3a outlines our experimental protocol, which is adapted from previous work²⁵. Briefly, pregnant dams were administered a total dose of 5 g kg^{-1} of 28% ethanol by intraperitoneal injection on E7.5. Control mice consisted of females from the same type of cross, subjected to injections of 0.9% saline solution. On E16.5, embryos were dissected and genotyped. The data in Fig. 3b and Supplementary Fig. 5d show that ethanol exposure reduces the frequency of double-mutant mice to 2.9% (7/240) compared to 14.5% (20/138) in the saline-treated group. A significant proportion of double-mutant embryos exposed *in utero* to ethanol suffered developmental defects, most drastically manifested as exencephaly (Fig. 3c–e, Supplementary Fig. 5f). The teratogenic events observed in embryos characterized in this experiment are compiled in Supplementary Fig. 5e. It is also noteworthy that a proportion of $\text{Aldh2}^{+/-}\text{Fancd2}^{-/-}$ embryos showed developmental defects in response to ethanol exposure (8/25; Fig. 3c, d and Supplementary Fig. 5e). In conclusion, $\text{Aldh2}^{-/-}\text{Fancd2}^{-/-}$ embryos are very sensitive to exposure to an unambiguous exogenous precursor of acetaldehyde.

Ethanol impairs haematopoiesis in $\text{Aldh2}^{-/-}\text{Fancd2}^{-/-}$ mice

Bone marrow failure is a hallmark of patients with Fanconi anaemia. It is also known that chronic ethanol abuse is associated with bone marrow dysfunction^{28,29} and that exposing bone marrow cells to ethanol or acetaldehyde compromises blood forming units³⁰. We therefore investigated the susceptibility of $\text{Aldh2}^{-/-}\text{Fancd2}^{-/-}$ mice to ethanol-induced myelotoxicity, the expectation being that ethanol exposure would lead to acetaldehyde accumulation and that this would then

cause bone marrow dysfunction. We exposed 6–8-week-old mice corresponding to four possible genotypes (wild type, $\text{Aldh2}^{-/-}$, $\text{Fancd2}^{-/-}$ and $\text{Aldh2}^{-/-}\text{Fancd2}^{-/-}$) to continuous oral ethanol in their drinking water supply. The protocol is outlined in Fig. 4a: it consists of 5 days of 15% ethanol/water followed by 5 days with 20% ethanol/water. The data in Supplementary Fig. 6 show that ethanol treatment causes a decrease in all three blood constituents. This is most notable in double-mutant mice where there is a profound drop in red blood cells and haemoglobin (Supplementary Fig. 6a, d). To establish the basis for this anaemia, we quantified the bone marrow cellularity in treated mice compared to untreated control animals. Figure 4b shows a significant decrease in the number of nucleated cells per femur in the ethanol-exposed double-mutant mice. To establish if DNA damage contributed to this ethanol sensitivity, bone marrow aspirates from ethanol-exposed mice were tested for the induction of phosphorylated histone γH2AX —an established marker of DNA damage. The western blot in Fig. 4c shows a clear induction of γH2AX in bone marrow cells obtained from double-mutant compared to control genotypes. We then set out to determine if progenitor colony-forming units (pre-B cells, granulocytes and erythrocytes) are affected by ethanol. Equal cell numbers of bone marrow cells were plated onto clonogenic plates, and it is clear that colony-forming units corresponding to all three cell types are reduced in double-mutant mice after ethanol treatment (Fig. 4d). Finally, we analysed the bone marrow histology, comparing untreated control mice with corresponding treated genotypes. The microscopy shown in Fig. 4e shows that ethanol consumption causes an almost complete obliteration of bone marrow haematopoiesis in $\text{Aldh2}^{-/-}\text{Fancd2}^{-/-}$ mice.

Leukaemogenesis in weaned double-mutant mice

Lastly, we wanted to determine the fate of the unexposed weaned $\text{Aldh2}^{-/-}\text{Fancd2}^{-/-}$ mice. Despite subtle developmental defects, double-mutant mice were nevertheless healthy. However, within 3–6 months a significant proportion of animals succumbed to an acute illness presenting itself as rapid weight loss and lethargy (Fig. 5a). These mice were killed and the majority of necropsies revealed a large mediastinal mass, splenomegaly and blast-like lymphoid cells in the peripheral blood film and bone marrow aspirate (Fig. 5b and Supplementary Fig. 7). In three instances, these cells represented a clonal expansion of early T cells as they were double positive for CD8 and CD4 markers and in one instance, CD8 alone (Supplementary Fig. 8a). Furthermore, immunohistochemistry revealed that the neoplastic cells strongly stained for the pan-T-cell-marker CD3 (Supplementary Fig. 8b). Because bone marrow aspirates indicated that >30% of the cellular population were blasts, we conclude that this illness is akin to acute lymphoblastic leukaemia.

Discussion

Children with Fanconi anaemia suffer developmental defects, bone marrow failure and cancer predisposition because of endogenous DNA damage². Aldehydes are likely to be a significant source of such damage, contributing to the genesis of the Fanconi anaemia clinical phenotype. Aldehyde accumulation is particularly toxic to Fanconi anaemia pathway-deficient bone marrow cells. Although the most obvious target is the haematopoietic stem cell pool, it is also possible that cells contributing to the marrow niche might also be susceptible. Our work also raises new therapeutic approaches to treat Fanconi anaemia in humans. For instance, it might be possible to induce the catabolism of aldehydes by induction of the relevant enzymes. This could be done simply—and somewhat paradoxically—by regular low-dose ethanol exposure or barbiturates^{31,32}. Alternatively, small-molecule agonists of Aldh2, like the Alda1 molecule, might raise acetaldehyde catabolic activity^{33,34}. Lastly, it will be important to determine the natural sources of aldehydes. These metabolic pathways could then be modulated so as to subvert toxic accumulation.

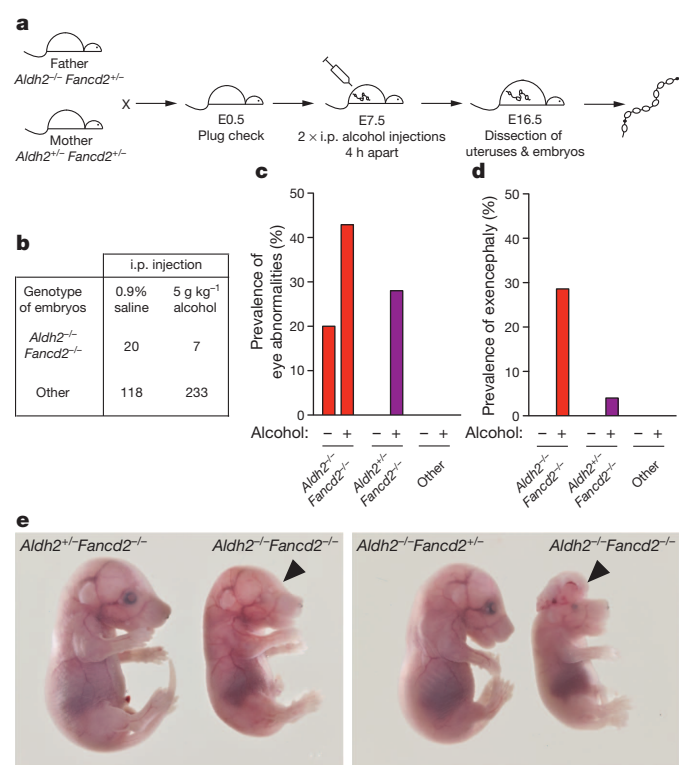


Figure 3 | Maternal ethanol exposure aborts the development of $\text{Aldh2}^{+/-}\text{Fancd2}^{-/-}$ embryos. **a**, Schematic representation of the experiment to expose pregnant $\text{Aldh2}^{+/-}\text{Fancd2}^{-/-}$ mice to ethanol. **b**, Numbers of $\text{Aldh2}^{+/-}\text{Fancd2}^{-/-}$ embryos after exposure to ethanol or saline solution. *** $P < 10^{-4}$, Fisher's exact test, confidence interval 1%. **c, d**, Prevalence of eye abnormalities (**c**) and exencephaly (**d**) for the various Aldh2 and Fancd2 genotypes in the saline (–) and ethanol (+) groups. **e**, Developmental abnormalities (left, anophthalmia; right, exencephaly).

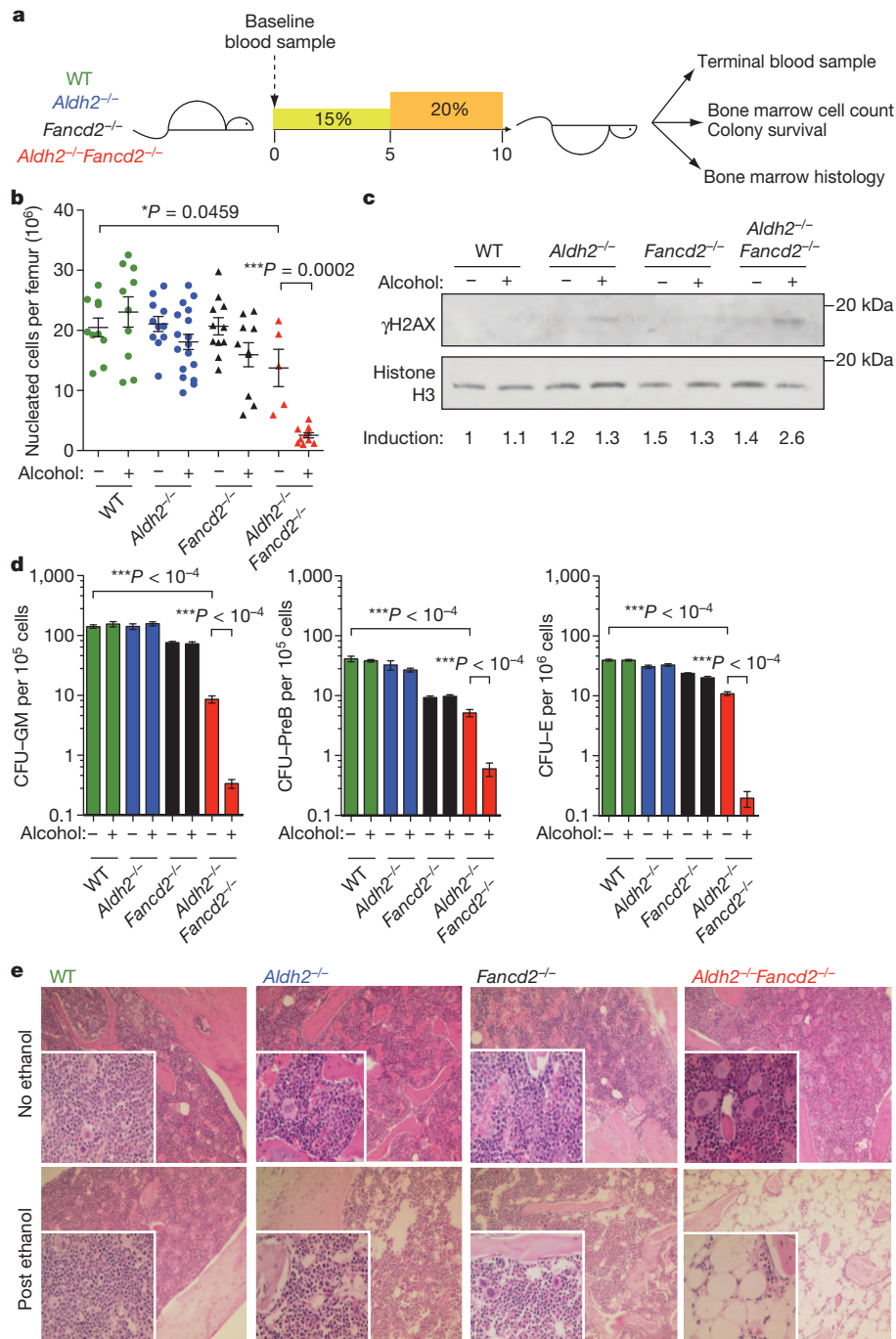


Figure 4 | Ethanol-induced bone marrow failure in *Aldh2*^{-/-}*Fancd2*^{-/-} mice. **a**, Schematic outline of the oral alcohol experiment. **b**, Bone marrow cellularity in untreated (-) and treated (+) mice. Unpaired *t*-test, central bar represents mean; error bars define s.e.m. **c**, Anti- γ H2AX western blot in untreated (-) and ethanol-exposed (+) bone marrow. **d**, Clonogenicity of untreated (-) and ethanol-exposed (+) bone marrow progenitors. Left: colony-forming-unit-granulocyte-monocyte cells (CFU-GM); middle: CFU-

pre-B cells; right: CFU-erythrocytes (CFU-E). Each bar represents the mean of three independent experiments, with two mice per experiment, each plated in duplicate; error bars define s.e.m. $^{***}P < 10^{-4}$, unpaired *t*-test.

e, Haematoxylin and eosin staining of bone marrow sections from mice not exposed (top) or exposed to ethanol (bottom). For each genotype/treatment unit, the lower magnification is $\times 100$, inset is $\times 400$.

Although it is currently unclear how aldehydes damage DNA *in vivo*, they directly modify bases *in vitro*, which can lead to DNA-protein and DNA-DNA crosslinks (Fig. 5c). *Aldh2* efficiently oxidizes acetaldehyde, but this enzyme also detoxifies other reactive aldehydes, such as 4-hydroxynonenal, acrolein, propionaldehyde and butyraldehyde³⁴. All these reactive molecules are generated through metabolism and may also contribute to DNA damage^{5,8}. It is also possible that aldehyde-mediated genotoxicity may be indirect; for example, it may deplete cellular NAD⁺ pools or stimulate the formation of other free

radical species. Nevertheless, our work indicates that the Fanconi anaemia pathway genes seem to be specifically required for cellular resistance to acetaldehyde. However, other genes working in homologous recombination (except *Rad51C*), and in translesion synthesis do not appear to confer cellular resistance to acetaldehyde. This is a surprise because genetic and biochemical work have identified a critical role for homologous recombination and translesion synthesis genes in DNA interstrand crosslink repair^{14,35}. It is therefore possible that aldehyde-mediated DNA damage does not result in DNA interstrand crosslinks.

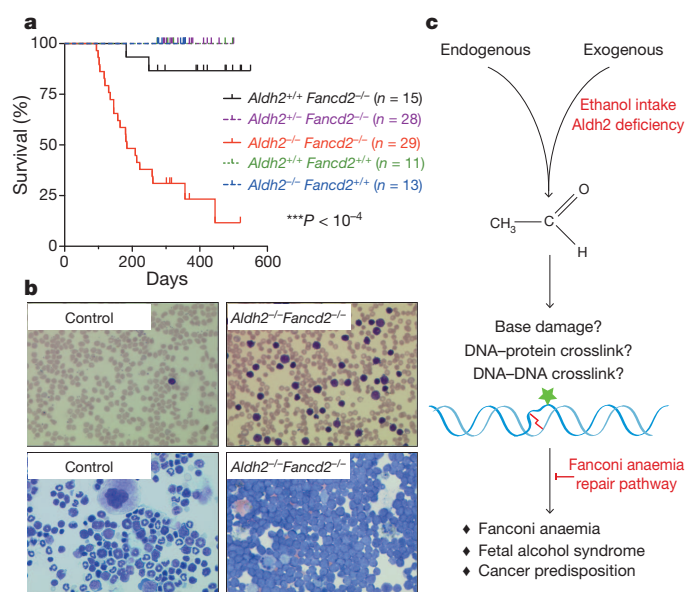


Figure 5 | Acute leukaemia in *Aldh2*^{-/-}*Fancd2*^{-/-} mice. **a**, Kaplan-Meier survival curve for the cohort of weaned *Aldh2*^{-/-}*Fancd2*^{-/-} mice and relevant genotype controls (black, purple, green and blue). ****P* < 10⁻⁴. **b**, Analysis of blood films (top; ×400) and bone marrow aspirates (bottom; ×400) obtained from a control (left) and a sick *Aldh2*^{-/-}*Fancd2*^{-/-} (right) mouse, revealing leukaemic blast cells. **c**, Cartoon summarizing the genetic interaction between acetaldehyde metabolism with DNA repair and the consequences of their dysfunction in humans.

In humans, excessive maternal alcohol consumption causes fetal alcohol syndrome, a prevalent worldwide cause of congenital learning disability³⁶. Our study potentially links fetal alcohol syndrome to DNA damage and therefore has far-reaching implications for the understanding of the pathogenesis of this affliction in humans. Although at present it is not known whether fetal alcohol syndrome is associated with cancer predisposition, recent epidemiological evidence indicates that maternal ethanol exposure may correlate with an increased risk of acute childhood leukaemia^{37,38}. The further relevance of this study to human health is also supported by the high frequency of *ALDH2* deficiency in South East Asians. In fact, up to 8% (540 million) of the world's population carries a dominant-negative mutation in *ALDH2* (ref. 39). Alcohol consumption in these individuals is strongly associated with a risk of aerodigestive tract cancer^{39,40} and, in the light of this study, it is very likely that acetaldehyde-mediated DNA damage drives cancer development. This might also apply to Fanconi anaemia patients, who have a high incidence of such cancers⁴¹. In summary, the findings reported here not only provide fresh insights into the interplay between metabolism and cancer predisposition, but also may have profound implications for the clinical care of Fanconi anaemia patients as well as overall public health.

METHODS SUMMARY

DT40 and mouse cells. Cells were grown in RPMI medium complemented with 3% chicken serum, 7% fetal calf serum, 50 μM β-mercaptoethanol and penicillin/streptomycin mix, at 37 °C. Homology arms for the *GgAldh2* targeting vector were amplified from DT40 genomic DNA and cloned in pBscript (Supplementary Fig. 1a). Sensitivity clonogenic assays were performed on methylcellulose after exposure for 2 h and 10–14 days incubation (see Methods). Lymphocytes purified from mouse spleens were stimulated with 40 μg ml⁻¹ LPS and plated with acetaldehyde. Viability was measured by Trypan blue exclusion after 7 days using a ViCell XR cell counter.

Mice. *Fancd2* mice were a gift from M. Grompe. *Aldh2*^{+/-} embryonic stem cells were injected in blastocysts from C57BL/6J mice. Chimaeric males were bred with C57BL/6J females to obtain germline transmission. Mice were maintained in specific pathogen-free conditions. All animal experiments undertaken in this study were done with the approval of the UK Home Office and the MRC Centre Ethical Review Committee.

Alcohol administration in mice. Pregnant females were injected intraperitoneally with 5 g kg⁻¹ of a 28% ethanol solution at E7.5. For the 10-day oral exposure experiment, water supply was replaced by a 15% and 20% ethanol solution and was ingested orally *ad libitum* by the mice.

Haematological studies. Peripheral blood counts were collected in EDTA tubes and analysed on a VetABC analyser. Haematopoietic colony forming unit assays were performed as described previously²¹ on bone marrow cells from mice exposed to alcohol for 10 days.

Histology. Tissue biopsies were fixed in 10% formalin, paraffin embedded and 4-μm sections were cut before haematoxylin and eosin staining. Blood smears and bone marrow cytopsins were stained with May–Grünwald and Giemsa stains.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 16 November 2010; accepted 11 May 2011.

- Lindahl, T. Instability and decay of the primary structure of DNA. *Nature* **362**, 709–715 (1993).
- Patel, K. J. & Joenje, H. Fanconi anemia and DNA replication repair. *DNA Repair (Amst.)* **6**, 885–890 (2007).
- O'Brien, P. J., Siraki, A. G. & Shangari, N. Aldehyde sources, metabolism, molecular toxicity mechanisms, and possible effects on human health. *Crit. Rev. Toxicol.* **35**, 609–662 (2005).
- Wang, M. *et al.* Identification of DNA adducts of acetaldehyde. *Chem. Res. Toxicol.* **13**, 1149–1157 (2000).
- Stein, S., Lao, Y., Yang, I. Y., Hecht, S. S. & Moriya, M. Genotoxicity of acetaldehyde- and crotonaldehyde-induced 1,N²-propanodeoxyguanosine DNA adducts in human cells. *Mutat. Res.* **608**, 1–7 (2006).
- Cheng, G. *et al.* Reactions of formaldehyde plus acetaldehyde with deoxyguanosine and DNA: formation of cyclic deoxyguanosine adducts and formaldehyde cross-links. *Chem. Res. Toxicol.* **16**, 145–152 (2003).
- Vasilou, V., Pappa, A. & Estey, T. Role of human aldehyde dehydrogenases in endobiotic and xenobiotic metabolism. *Drug Metab. Rev.* **36**, 279–299 (2004).
- Perez-Miller, S. *et al.* Alda-1 is an agonist and structural chaperone for the common human aldehyde dehydrogenase 2 variant. *Nature Struct. Mol. Biol.* **17**, 159–164 (2010).
- Ridpath, J. R. *et al.* Cells deficient in the FANCD2/BRCA pathway are hypersensitive to plasma levels of formaldehyde. *Cancer Res.* **67**, 11117–11122 (2007).
- Nagayoshi, H. *et al.* Increased formation of gastric N²-ethylidene-2'-deoxyguanosine DNA adducts in aldehyde dehydrogenase-2 knockout mice treated with ethanol. *Mutat. Res.* **673**, 74–77 (2009).
- Matsuda, T. *et al.* Increased formation of hepatic N²-ethylidene-2'-deoxyguanosine DNA adducts in aldehyde dehydrogenase-2 knockout mice treated with ethanol. *Carcinogenesis* **28**, 2363–2366 (2007).
- Seitz, H. K. & Stickel, F. Molecular mechanisms of alcohol-mediated carcinogenesis. *Nature Rev. Cancer* **7**, 599–612 (2007).
- Chen, L. *et al.* Quantitation of an acetaldehyde adduct in human leukocyte DNA and the effect of smoking cessation. *Chem. Res. Toxicol.* **20**, 108–113 (2007).
- Niedzwiedz, W. *et al.* The Fanconi anaemia gene FANCD2 promotes homologous recombination and error-prone DNA repair. *Mol. Cell* **15**, 607–620 (2004).
- Alpi, A. *et al.* UBE2T, the Fanconi anaemia core complex, and FANCD2 are recruited independently to chromatin: a basis for the regulation of FANCD2 monoubiquitination. *Mol. Cell Biol.* **27**, 8421–8430 (2007).
- King, G. & Holmes, R. Human ocular aldehyde dehydrogenase isozymes: distribution and properties as major soluble proteins in cornea and lens. *J. Exp. Zool.* **282**, 12–17 (1998).
- Pappa, A., Estey, T., Manzer, R., Brown, D. & Vasilou, V. Human aldehyde dehydrogenase 3A1 (ALDH3A1): biochemical characterization and immunohistochemical localization in the cornea. *Biochem. J.* **376**, 615–623 (2003).
- Riveros-Rosas, H., Julian-Sanchez, A. & Pina, E. Enzymology of ethanol and acetaldehyde metabolism in mammals. *Arch. Med. Res.* **28**, 453–471 (1997).
- Kunitoh, S. *et al.* Acetaldehyde as well as ethanol is metabolized by human CYP2E1. *J. Pharmacol. Exp. Ther.* **280**, 527–532 (1997).
- Parmar, K., D'Andrea, A. & Niedernhofer, L. J. Mouse models of Fanconi anemia. *Mutat. Res.* **668**, 133–140 (2009).
- Crossan, G. P. *et al.* Disruption of mouse Slx4, a regulator of structure-specific nucleases, phenocopies Fanconi anemia. *Nature Genet.* **43**, 147–152 (2011).
- Marietta, C., Thompson, L. H., Lamerdin, J. E. & Brooks, P. J. Acetaldehyde stimulates FANCD2 monoubiquitination, H2AX phosphorylation, and BRCA1 phosphorylation in human cells *in vitro*: implications for alcohol-related carcinogenesis. *Mutat. Res.* **664**, 77–83 (2009).
- Yu, H. S. *et al.* Characteristics of aldehyde dehydrogenase 2 (*Aldh2*) knockout mice. *Toxicol. Mech. Methods* **19**, 535–540 (2009).
- Houghtaling, S. *et al.* Epithelial cancer in Fanconi anemia complementation group D2 (*Fancd2*) knockout mice. *Genes Dev.* **17**, 2021–2035 (2003).
- Sulik, K. K., Johnston, M. C. & Webb, M. A. Fetal alcohol syndrome: embryogenesis in a mouse model. *Science* **214**, 936–938 (1981).
- Webster, W. S., Walsh, D. A., McEwen, S. E. & Lipson, A. H. Some teratogenic properties of ethanol and acetaldehyde in C57BL/6J mice: implications for the study of the fetal alcohol syndrome. *Teratology* **27**, 231–243 (1983).
- O'Shea, K. S. & Kaufman, M. H. The teratogenic effect of acetaldehyde: implications for the study of the fetal alcohol syndrome. *J. Anat.* **128**, 65–76 (1979).

28. Michot, F. & Gut, J. Alcohol-induced bone marrow damage. A bone marrow study in alcohol-dependent individuals. *Acta Haematol.* **78**, 252–257 (1987).
29. Nakao, S., Harada, M., Kondo, K., Mizushima, N. & Matsuda, T. Reversible bone marrow hypoplasia induced by alcohol. *Am. J. Hematol.* **37**, 120–123 (1991).
30. Meagher, R. C., Sieber, F. & Spivak, J. L. Suppression of hematopoietic-progenitor-cell proliferation by ethanol and acetaldehyde. *N. Engl. J. Med.* **307**, 845–849 (1982).
31. Marc, N., Fautrel, A., Damon, M., Guillozo, A. & Corcos, L. Phenobarbital induction of aldehyde dehydrogenase type 2 mRNA in mouse liver: a candidate region on chromosome 7 for a putative regulatory gene. *Biochem. Genet.* **38**, 297–308 (2000).
32. Vasiliou, V., Torronen, R., Malamas, M. & Marselos, M. Inducibility of liver cytosolic aldehyde dehydrogenase activity in various animal species. *Comp. Biochem. Physiol. C* **94**, 671–675 (1989).
33. Chen, C.-H. *et al.* Activation of aldehyde dehydrogenase-2 reduces ischemic damage to the heart. *Science* **321**, 1493–1495 (2008).
34. Perez-Miller, S. *et al.* Alda-1 is an agonist and chemical chaperone for the common human aldehyde dehydrogenase 2 variant. *Nature Struct. Mol. Biol.* **17**, 159–164 (2010).
35. Knipscheer, P. *et al.* The Fanconi anemia pathway promotes replication-dependent DNA interstrand cross-link repair. *Science* **326**, 1698–1701 (2009).
36. Abel, E. L. & Sokol, R. J. A revised conservative estimate of the incidence of FAS and its economic impact. *Alcohol. Clin. Exp. Res.* **15**, 514–524 (1991).
37. Latino-Martel, P. *et al.* Maternal alcohol consumption during pregnancy and risk of childhood leukemia: systematic review and meta-analysis. *Cancer Epidemiol. Biomarkers Prev.* **19**, 1238–1260 (2010).
38. MacArthur, A. C. *et al.* Risk of childhood leukemia associated with parental smoking and alcohol consumption prior to conception and during pregnancy: the cross-Canada childhood leukemia study. *Cancer Causes Control* **19**, 283–295 (2008).
39. Brooks, P. J., Enoch, M. A., Goldman, D., Li, T. K. & Yokoyama, A. The alcohol flushing response: an unrecognized risk factor for esophageal cancer from alcohol consumption. *PLoS Med.* **6**, e50 (2009).
40. McKay, J. D. *et al.* A genome-wide association study of upper aerodigestive tract cancers conducted within the INHANCE Consortium. *PLoS Genet.* **7**, e1001333 (2011).
41. Alter, B. P., Joenje, H., Oostra, A. B. & Pals, G. Fanconi anemia: adult head and neck cancer and hematopoietic mosaicism. *Arch. Otolaryngol. Head Neck Surg.* **131**, 635–639 (2005).
42. Hazen, S. L., Hsu, F. F., d'Avignon, A. & Heinecke, J. W. Human neutrophils employ myeloperoxidase to convert α -amino acids to a battery of reactive aldehydes: a pathway for aldehyde generation at sites of inflammation. *Biochemistry* **37**, 6864–6873 (1998).
43. O'Brien, P. J., Siraki, A. G. & Shangari, N. Aldehyde sources, metabolism, molecular toxicity mechanisms, and possible effects on human health. *Crit. Rev. Toxicol.* **35**, 609–662 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Grompe for *Fancd2* knockout mice, J. Sale and S. Takeda for DT40 strains, N. Sugimura and F. Gergely for comments on the manuscript. We are grateful to T. Langford, R. Berks, V. Smith, J. Wiles, C. Shepherd, M. Kidd, M. Brown, A. Mead, R. Pannell, J. Garaycochea and A. Shortland for their assistance with animal experiments and husbandry. We thank N. Grant and P. Banks from the Visual Aids department for photographic images and W. Zhao of the Human Research Tissue Bank (NIHR Cambridge Biomedical Research Centre) for processing histology. F.L. and I.V.R. are funded by the Children's Leukaemia Trust and Fanconi Anaemia Research Fund, respectively. K.J.P. acknowledges M. Neuberger, N. McIntyre and C. Desai for support.

Author Contributions K.J.P., F.L. and G.P.C. designed the experiments. F.L. and G.P.C. performed the majority of the experimental work, I.V.R. contributed to DT40 clonogenic assays and FACS analysis of tumours. M.J.A. carried out histological analysis. K.J.P. wrote the manuscript assisted by F.L. and G.P.C.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to K.J.P. (kjp@mrc-lmb.cam.ac.uk).

METHODS

DT40 cell culture and acetaldehyde sensitivity assays. DT40 cells were maintained in RPMI medium supplemented with 50 μ M β -mercaptoethanol, 7% fetal calf serum, 3% chicken serum and penicillin/streptomycin mix (Gibco) in a 10% CO₂ incubator at 37 °C. Sensitivity to acetaldehyde (Fluka; catalogue no. 00070) was measured by colony survival assays. 2×10^5 cells were treated with acetaldehyde for 2 h at 37 °C before seeding of 3 dilutions on methylcellulose-DMEM medium in 6-well plates. After 10–14 days incubation, colonies were scored and survival was plotted relative to the untreated control. Each experiment represents the average of 3 independent experiments.

Generation of Δ ALDH2 cells. Δ ALDH2 construct: the 5' homology arm was amplified from DT40 genomic DNA and cloned as a 2.7-kb XhoI-BamHI fragment in pBluescript (aldh2_5F1, 5'-TTGAACTCTGGATAATGGCAAC; and aldh2_5R1, 5'-AGCTTTCTTGATCAGGTGCCCAAC). The 3' homology arm was amplified from genomic DNA and cloned as a 2.4-kb BamHI-NotI fragment in pBluescript (aldh2_3F1, 5'-ATAGCCTATGTCTGCTTGGATCCACA; and aldh2_3R1, 5'-TGACGTCTGAAGAGTGCAGCTCCT). Before transfection, the targeting construct was linearized by NotI restriction digest. Drug-resistant clones were screened by Southern blot using a 900-bp fragment as a probe amplified from genomic DNA (aldh2sbpBF1, 5'-ATACCGATTCCTCAAGGT TTGGAT; and aldh2sbpBR1, 5'-CCACTATGTCATCTGACAGGTTGA).

Mouse strains. *Fancd2* mice (*Fancd2*^{tm1Hou}, MGI code: 2673422, 129S4/SvJae) were a gift from M. Grompe. *Aldh2* mice were generated from embryonic stem cells purchased from EUComm (*Aldh2*^{tm1a(EUComm)Wtsi}, MGI code: 4431566, C57BL/6N). Details of the targeting construct can be found in Supplementary Fig. 3 and <http://www.knockoutmouse.org/about/targeting-strategies>. Embryonic stem cells were injected in blastocyst to generate chimaeric animals. Germline transmission of the targeted allele was monitored by PCR using *Aldh2* forward (5'-TGGACATGGTCCTGAAATGTCTCC-3') and *Aldh2* reverse (5'-GACT AGACTGCCAGAAACCATGAA-3'); 95 °C for 3 min, 95 °C for 30 s, 65 °C for 30 s, 72 °C for 30 s (35 cycles), 72 °C for 3 min. In individual experiments, all mice were matched for age and gender. Mice were maintained in specific pathogen-free conditions. All animal experiments undertaken in this study were done with the approval of the UK Home Office and the MRC Centre Ethical Review Committee.

Pregnancy experiment. Timed matings of *Aldh2*^{-/-}*Fancd2*^{+/-} males and *Aldh2*^{+/-}*Fancd2*^{+/-} females were set up. Females were checked for the presence of a vaginal plug the following morning, considered as day E0.5. Plugged females were injected with either 5 g kg⁻¹ of a 28% ethanol solution or 0.9% saline

solution equivalent at E7.5, in two separate intraperitoneal injections of 2.5 g kg⁻¹, 4 h apart. At E16.5, pregnant females were killed and uteruses taken for dissection of embryos.

Oral alcohol experiment. Six-to-eight-week-old mice of four possible genotypes (wild type, *Aldh2*^{-/-}, *Fancd2*^{-/-} and *Aldh2*^{-/-}*Fancd2*^{-/-}) were exposed to continuous oral alcohol exposure for 10 days. For the first 5 days, the drinking water supply was replaced by a 15% ethanol/water solution, followed by a 20% ethanol/water solution for the last 5 days. A baseline blood sample was taken from tails before alcohol exposure and by cardiac puncture at the end of the experiment, in order to measure full blood counts. Femurs were dissected for histological analysis and to determine bone marrow cellularity. Bone marrow of individual femurs was flushed out in 400 μ l of PBS using a 26-gauge needle. Nucleated cells were quantified using 3% acetic acid and methylene blue and a ViCell XR counter (Beckman Coulter).

Peripheral blood counts. Blood was collected in EDTA microvette tubes (Sarstedt) and analysed on a VetABC analyser.

Haematopoietic CFU assays. These were performed using total bone marrow cells harvested from the femurs and tibias of untreated mice or after 10 days of oral alcohol exposure. Nucleated cells were enumerated using 3% acetic acid and methylene blue. 10^7 , 10^6 and 10^5 nucleated cells were plated in Methocult M3334 (Stem Cell Technologies) and CFU-E was counted after 9 days. 10^6 , 10^5 , and 10^4 nucleated cells were plated in either Methocult GF M3534 (Stem Cell Technologies) for CFU-GM or Methocult M3630 (Stem Cell Technologies) for CFU-pre-B and counted after 7 days²¹.

Sensitivity assays of primary mouse B cells. These were performed on lymphocytes purified from the spleen using Lympholyte M (Cederlane). Lymphocytes were stimulated with LPS (Sigma L4391) at a final concentration of 40 μ g ml⁻¹. 4×10^5 cells were plated with acetaldehyde in one well of a 24-well plate. After 7 days the viable cells were enumerated using trypan blue exclusion counting 100 images using a ViCell XR (Beckman Coulter). Each data point represents the mean of two independent experiments each carried out in triplicate.

Immunoblot. γ H2A.X antibody (Millipore, JBW301) was used at 1:3,000. Histone H3 polyclonal antibody (Abcam) was used at 1:2,000. FANCD2 anti-serum was used at 1:3,000. Fold-induction of γ H2A.X relative to H3 was calculated by densitometry.

Histology. Tissue biopsies were paraffin embedded and 4- μ m sections were cut before haematoxylin and eosin staining. Blood smears and bone marrow cytopspins were stained with May-Grünwald and Giemsa stains.

Crystal structure of a copper-transporting PIB-type ATPase

Pontus Gourdon^{1*}, Xiang-Yu Liu^{1,2*}, Tina Skjørringe³, J. Preben Morth^{1†}, Lisbeth Birk Møller³, Bjørn Panyella Pedersen^{1†} & Poul Nissen¹

Heavy-metal homeostasis and detoxification is crucial for cell viability. P-type ATPases of the class IB (PIB) are essential in these processes, actively extruding heavy metals from the cytoplasm of cells. Here we present the structure of a PIB-ATPase, a *Legionella pneumophila* CopA Cu⁺-ATPase, in a copper-free form, as determined by X-ray crystallography at 3.2 Å resolution. The structure indicates a three-stage copper transport pathway involving several conserved residues. A PIB-specific transmembrane helix kinks at a double-glycine motif displaying an amphipathic helix that lines a putative copper entry point at the intracellular interface. Comparisons to Ca²⁺-ATPase suggest an ATPase-coupled copper release mechanism from the binding sites in the membrane via an extracellular exit site. The structure also provides a framework to analyse missense mutations in the human ATP7A and ATP7B proteins associated with Menkes' and Wilson's diseases.

P-type ATPases are integral membrane pumps that derive energy from ATP hydrolysis to maintain ion homeostasis, electrochemical gradients and lipid bilayer asymmetry in cells^{1,2}. The P-type ATPase superfamily encompasses 11 distinct classes^{3,4}, of which class IB and IIA are the largest and most widespread, ranging from bacteria to humans³. Atomic structures have been determined for the class IIA sarcoplasmic reticulum Ca²⁺-ATPase (SERCA1a) in several conformations^{5–9}, the class IIC Na⁺,K⁺-ATPase^{10,11}, and the class IIIA H⁺-ATPase¹². The catalytic mechanism of P-type ATPases is described by the Albers–Post model^{13,14} with E1 and E2 states associated with high and low affinity, respectively, for the extruded substrate, and E1P and E2P denoting phosphoenzyme intermediates. The translocation of ions is facilitated by conformational changes coupled to ATP phosphorylation and subsequent dephosphorylation, controlling alternating access to transport sites in the membrane².

PIB-ATPases are heavy-metal-transporting ATPases that are essential cellular regulators of, for example, Cu⁺, Zn²⁺ and Co²⁺ (ref. 15). Tight regulation coupled to active transport is an essential process due to the toxicity of these metals, of which some are essential cofactors. Reflecting the extremely low intracellular concentrations of these metals^{16,17}, the PIB-ATPases are characterized by high apparent affinity (with dissociation constants in the femtomolar range¹⁸). A representative intramembranous CPX/XPC motif (typically CPC)^{19,20} is associated with heavy-metal binding. PIB-ATPases are predicted to share a core with other P-type ATPases based on three cytoplasmic domains as well as a transmembrane domain with six helices corresponding to the first six of SERCA1a (ref. 1). This core structure is frequently flanked by two additional, amino-terminal transmembrane helices and one or sequential heavy-metal binding domains (HMBD)¹⁵ with CXXC motif(s). It is not clear how and when these domains interact with the catalytic core or if and how they participate in sub-cellular targeting, regulation and/or ion transfer^{17,21,22}.

Cu⁺-transporting ATPases are the most prevalent PIB-ATPases¹⁵. In plants and many microorganisms they have an important role in copper homeostasis and detoxification^{23,24}. The two Cu⁺-transporting

PIB-ATPases in human are ATP7A and ATP7B, which are of vital importance^{25–27}; defects give rise to Menkes' and Wilson's diseases, respectively, whereas upregulation has been associated with Alzheimer's disease²⁸ and resistance to cancer chemotherapy²⁹.

Atomic structures of the cytoplasmic domains of PIB-ATPases have revealed class-specific features^{30–32}, but electron microscopy studies of full-length proteins have generated contrasting topology models^{33,34}. No atomic structure of a complete PIB-ATPase has been determined until now, and key questions on the enzyme structure—including how copper transport is coupled to the ATPase activity, how cytoplasmic copper is loaded and released from the protein, and how disease-related mutations are distributed in the protein—remain to be answered¹⁸.

Overall structure of CopA

The Cu⁺-ATPase structure described here is derived from the *L. pneumophila* gene *lpg1024*, which can complement a CopA-deficient strain of *Escherichia coli*³⁵. The Lpg1024 protein shows significant sequence identity to human ATP7A and ATP7B and other Cu⁺-conducting PIB-ATPases (Supplementary Fig. 1). We observe Cu⁺-dependent ATPase activity *in vitro* (Supplementary Fig. 2) and will refer to this protein as CopA or LpCopA (when specifically referring to the *L. pneumophila* protein). LpCopA was crystallized with high amounts of exogenous lipid, and the structure was determined from electron density maps with experimental phases obtained from multiple isomorphous replacement with anomalous scattering (MIRAS) and density modification (Supplementary Fig. 3 and Supplementary Table 1). The final model includes residues Val 74 to the carboxy-terminal Leu 736. The unmodelled 73 N-terminal residues encompass a single HMBD, which is, however, partially visible as low-resolution electron density features.

The cytosolic part of CopA displays three domains that are characteristic of P-type ATPases³⁶ (Fig. 1): the A-domain (actuator), P-domain (phosphorylation) and the N-domain (nucleotide binding). These are generally smaller than for other classes of P-type

¹Centre for Membrane Pumps in Cells and Disease—PUMPKIN, Danish National Research Foundation, Aarhus University, Department of Molecular Biology, Gustav Wieds Vej 10C, DK-8000 Aarhus C, Denmark. ²State Key Laboratory of Protein and Plant Gene Research, College of Life Sciences, Peking University, Beijing, 100871, China. ³Center for Applied Human Molecular Genetics, Kennedy Center, Gl. Landevej 7, 2600 Glostrup, Denmark. [†]Present addresses: The Biotechnology Centre of Oslo and Centre for Molecular Medicine, Nordic EMBL Partnership, University of Oslo, 0318 Oslo, Norway (J.P.M.); Department of Biochemistry and Biophysics, University of California at San Francisco, San Francisco, California 94158, USA (B.P.P.).

*These authors contributed equally to this work.

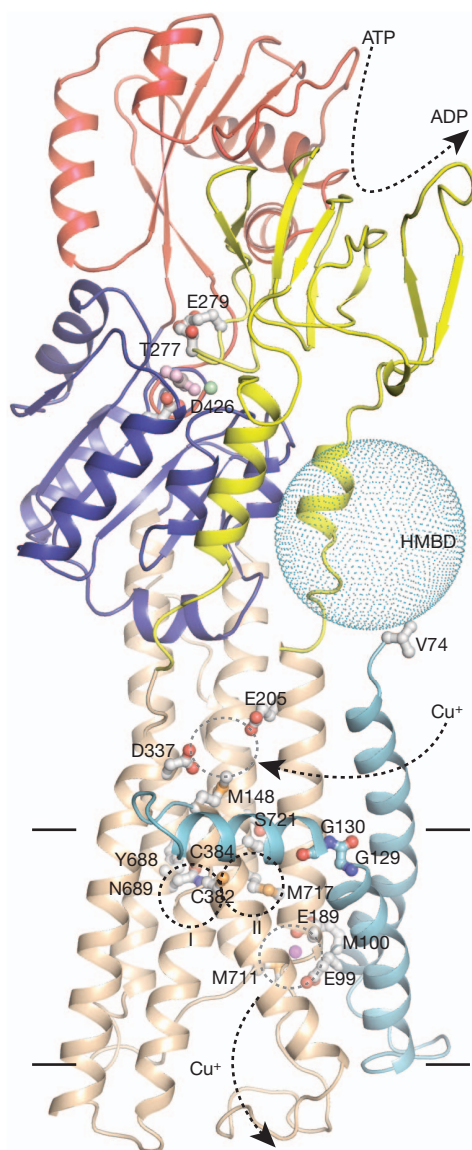


Figure 1 | Overall structure of the *L. pneumophila* Cu^+ -ATPase LpCopA. Cartoon representation of CopA with A-, N- and P-domains in yellow, red and blue, respectively. The transmembrane helices are displayed in cyan (MA–MB) and wheat (M1–M6). A dotted, cyan sphere indicates the HMBD region. AlF_4^- at Asp 426 and Glu 279 is pink and brown, and Mg^{2+} is in green. A K^+ ion modelled at the putative exit site is shown in purple. Key residues are shown in ball-and-stick representation. Arrows indicate the suggested Cu^+ transport pathway and ATP turnover, lines the approximate position of the membrane, whereas circles depict putative Cu^+ -binding sites at entry (grey), membrane (black) and exit (grey).

ATPases, but despite a low degree of sequence conservation (Supplementary Fig. 4) their core structures are conserved, as noted previously^{30–32}.

The M-domain consists of eight transmembrane segments: the six core helices with a principal organization as observed for the first six helices of other P-type ATPases^{5,10,12} (Fig. 2 and Supplementary Fig. 5) preceded by the two PIB-specific helices (MA and MB). This topology, and in particular the position of the latter two helices, has been a matter of debate^{1,20,33,34,37}. Furthermore, the A-domain comprises only the cytoplasmic loop between M2 and M3, lacking the N-terminal part. MA interacts with M2 and M6, and MB with M1 and M2. MB consists of a short transmembrane helix followed by a kink and an amphipathic helix MB' positioned at the cytoplasmic membrane interface (Figs 3

and 4a and Supplementary Fig. 6). We note that the M1 helix of CopA is not kinked as observed in other P-type ATPase structures^{6–12}.

LpCopA was crystallized in the absence of copper and in complex with AlF_4^- , representing an occluded, copper-released E2- P_i (where P_i indicates inorganic phosphate) transition state (see Supplementary Fig. 7). Comparison to SERCA1a structures shows the closest resemblance to the equivalent, proton-occluded E2- P_i transition state (root mean squared deviation (r.m.s.d.) 3.3 Å for Protein Data Bank accession 3b9r (ref. 9)), although CopA adopts a more compact configuration with M4 (including the CPC motif) shifted towards M1/M2 (Supplementary Fig. 5b).

The membrane copper-binding sites

The membranous ion-binding sites in P-type ATPases are typically denoted I and II, with site II being accessible through an N-terminal, cytoplasmic entrance pathway⁵ (Figs 1 and 2). Mutagenesis in conjunction with biochemical studies have indicated that six invariant residues in M4, M5 and M6 of CopA contribute to two ion-binding sites in the M-domain, but the exact stoichiometry of the transport reaction cycle (probably two Cu^+ per ATP) remains elusive¹⁸. Five of the invariant residues—Cys 384 (last of the CPC motif in M4), Tyr 688 and Asn 689 of M5, and Met 717 and Ser 721 of M6—overlap with the calcium-coordinating residues of SERCA1a in the calcium-free state (Fig. 2). In addition, Cys 382 replaces Ile 307 in SERCA1a providing a main-chain oxygen for calcium coordination at site II. Assuming that similar conformational changes occur in CopA, the copper-bound state will bear resemblance to the calcium-bound state of SERCA1a. Only Tyr 688 and Asn 689 of M5 would require side-chain rearrangements to reach the putative sites II and I, possibly assisted by Pro 694 (conserved in M5 of PIB-ATPases) as a helix break point.

Except for the generally conserved Pro 383 of M4 (Pro 308 in SERCA1a), the residues involved in ion binding in class II ATPases are not preserved in the class IB Cu^+ -ATPases. Furthermore, the only conserved and charged residue in the M-domain of CopA is Glu 189 in M2. This indicates that CopA may operate without counter transport, which indeed has never been demonstrated for a PIB-ATPase.

The heavy-metal binding domain

In LpCopA, an N-terminal Cu^+ -binding domain (a HMBD) appears before the MA helix. We, and others, have shown that it is not strictly required for CopA ATPase activity¹⁷ (Supplementary Fig. 2b, d). Structural studies of HMBDs^{38,39} and cytosolic copper chaperones^{40,41} have revealed a $\beta\alpha\beta\beta\alpha\beta$ fold, and with the known position of Val 74, this structural motif significantly limits the plausible position(s) for the domain.

Although N-terminal sequencing shows that LpCopA is intact, we were unable to model the HMBD due to weak or absent electron density for the region. However, data from several crystals revealed continuous electron density peripheral to the A-domain coinciding with an unassigned selenomethionine peak as well as a mercury site (presumably at the CXXC motif), about 21 Å from Val 74 (Fig. 3 and Supplementary Figs 3, 6, 8 and 9). Such a proposed HMBD position is corroborated by biochemical data³⁷ and coincides with one of three positions suggested by electron microscopy³⁴.

The proposed HMBD location coincides with the linker from the N-terminal part of the A-domain to M1 in SERCA1a—a feature missing in PIB-ATPases (Supplementary Fig. 10). The integrity of this linker is essential for the functional cycle of SERCA1a (ref. 42), suggesting that the HMBD might regulate CopA function through interactions with the A-domain, as also indicated from biochemical studies³⁷. However, further analysis of the crystal structure as well as the electron microscopy data³⁴ and biochemical data⁴³ indicate that the HMBD may also interact at other positions, possibly exploiting both inhibitory and/or stimulatory functions in CopA regulation (Supplementary Fig. 6).

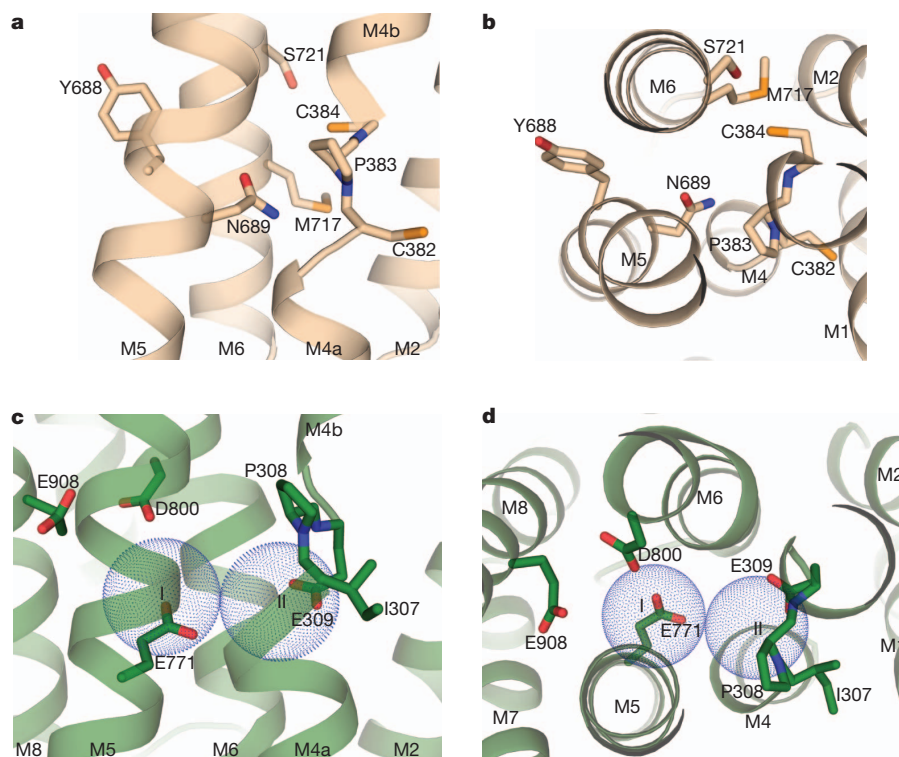


Figure 2 | Details of CopA compared to the binding sites I and II in SERCA1a in the equivalent calcium-free state. For CopA, transmembrane helices M1–M6 are shown in wheat, whereas SERCA1a is displayed in green (PDB 3b9r (ref. 9)). **a**, View in the plane of the membrane of the transmembrane domain showing the residues associated with ion-binding sites

of CopA¹⁸. Site I: Asn 689 (M5), Met 717 (M6) and Ser 721 (M6). Site II: Cys 382 and Cys 384 (CPC motif of M4) as well as Tyr 688 (M5). **b**, Same residues as in **a**, seen from the cytoplasmic side. **c**, Equivalent view for SERCA1a as in **a**. The approximate positions of sites I and II (in the E1 states) are indicated by blue dotted spheres. **d**, Equivalent view as in **b** for SERCA1a.

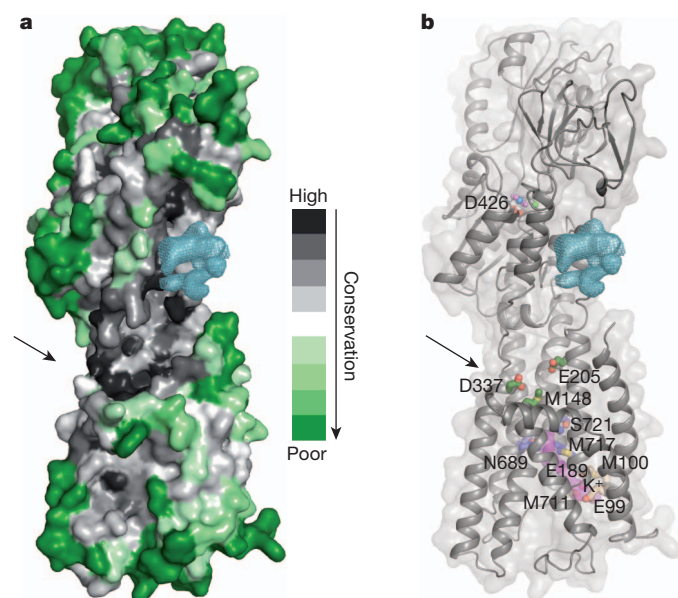


Figure 3 | The cytoplasmic platform and the heavy-metal binding domain. Electron density (blue mesh) adjacent to the A-domain marks a putative position of the HMBD. Arrows indicate the amphipathic MB' helix. See also Supplementary Fig. 6. **a**, The surface conservation among CopA proteins (alignment in Supplementary Fig. 1a) is depicted in a black–green colour ramp highlighting the proposed entry site. **b**, Cartoon representation with residues of the proposed Cu²⁺ binding sites and phosphorylation site Asp 426 indicated in stick representation. A putative exit pathway (computed using the software CAVER⁴⁹, shown in purple) leads from Cys 384 and Met 717 through the putative exit site and out. Entry site residues are shown in green stick representation, binding site in blue, and exit site in wheat.

A putative copper transport pathway

The class-IB-specific MA/MB transmembrane helices are of special interest: Val 74 initiates the long and curved MA helix, and the N-terminal part of MB provides a short transmembrane helix that kinks at the cytosolic membrane interface (facilitated by two conserved glycines, Gly 129 and Gly 130; Fig. 1). The C-terminal part of MB forms an amphipathic MB' helix with bulky, hydrophobic residues directed towards the membrane and positively charged residues facing the cytoplasm (Fig. 4a). The consensus sequence of CopA proteins preserves the GG kink and the highly amphipathic nature of MB' (Supplementary Fig. 1).

MB' and M1 form a platform (Figs 3 and 4a), which, by analogy to SERCA1a, may be part of the Cu²⁺ entry pathway (Fig. 2c, d)^{6,44}. Examination of conserved residues exposed at the platform unveils three candidate residues for initial Cu²⁺ coordination: Met 148, Glu 205 and Asp 337 (Fig. 4a). Met 148 and Asp 337 are ~5 Å from each other and Glu 205 is ~5.5 and ~7.5 Å from them. The putative entry site is located ~23 Å from Val 74, similar to the distance from Val 74 to the mercury site expected to mark the CXXC motif of the HMBD.

We believe that this platform provides a docking site for the HMBD or a soluble copper chaperone, either for copper delivery²² and/or for CopA auto-regulation by the HMBD¹⁷. The positively charged residues surrounding the putative entry site (Fig. 4a and Supplementary Fig. 6) comply with 'the positive-inside' rule⁴⁵ and may steer attraction or repulsion of the HMBD or soluble copper chaperones by electrostatic interactions, as complex formation between a human HMBD and a soluble chaperone shows charge complementation at the interacting surfaces⁴⁶.

We speculate that Cu²⁺ is passed from the HMBD or a soluble copper chaperone to the platform entry site, conceivably during the E2 to E1 transition (Figs 1, 3b and 4a). For translocation to the

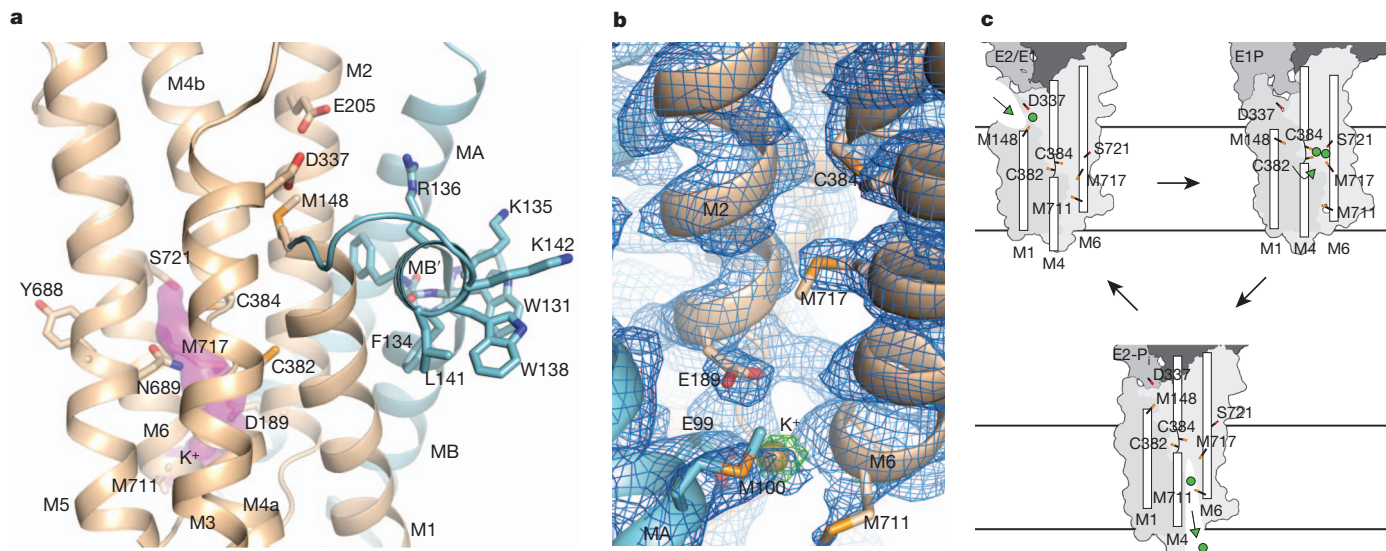


Figure 4 | Proposed stages at the copper transport pathway of CopA. The domains of CopA are coloured as in Fig. 1. **a**, A platform at the cytoplasmic interface is defined by the amphipathic MB' helix and displays an entry site (Met 148, Glu 205 and Asp 337) leading to the CPC motif in M4. A putative exit pathway is illustrated as in Fig. 3b. **b**, The putative Cu^+ exit site of CopA. An ion, assigned as K^+ , binds at an extracellular site (beige sphere) defined by the conserved Glu 189 and Met 100 and the less conserved Glu 99 and Met 711.

membrane site II, the CPC motif may be of significance. Cys 382 is oriented towards the cytoplasm with an S–S distance to Met 148 of ~ 9.5 Å (Fig. 4a). In SERCA1a, the E2 to E1 transition is associated with a movement of M1 towards the extracellular side and M4 towards the cytoplasm—these conformational changes are likely to be universal modes of P-type ATPase function. Therefore, we predict a significantly increased distance between Met 148, Glu 205 and Asp 337, and a reduced distance between Met 148 and Cys 382 as a result of the E2 to E1 transition of CopA. Such alterations may assist in Cu^+ transfer from the entry site at the platform to site II in the membrane (Fig. 4a, c and Supplementary Fig. 11). The simultaneous rotational shift of M4 (known for SERCA1a) and a switch in conformation at the CPC motif could assist in establishing the occluded, high-affinity transmembrane ion-binding sites I and II.

The relatively high apparent Cu^+ affinity for CopA proteins raises the question of how Cu^+ is released. Our structure, and specifically the way that Cys 382 is oriented, provides insight (Fig. 4c and Supplementary Fig. 11): space is permitted by the conserved Gly 155 to let the Cys 382 side chain make a snug fit against M1 and away from site II, stabilized by hydrogen bonds to the backbone carbonyls of Leu 151 (~ 2.9 Å) and possibly Ile 152 (~ 3.5 Å). This buried and stabilized position, coupled to the dephosphorylation reaction, excludes the Cys 382 side chain from copper binding and probably distorts the high-affinity transmembrane Cu^+ sites, stimulating Cu^+ release to the extracellular side.

We suggest that Cu^+ release is further stimulated by the conserved and negatively charged residue Glu 189 (Fig. 4b, c). The equivalent residue in SERCA1a (Glu 90) serves such a purpose for calcium release⁹. Significant residual electron density (at 5.1σ in the $F_o - F_c$ map before site modelling) is found at a site adjacent to Glu 189 that includes Met 100 and Met 711 (conserved in CopA) as well as Glu 99 (less conserved). We have modelled this putative Cu^+ exit site occupant as a K^+ ion. As for the entry site (Met 148/Glu 205/Asp 337), the combination of carboxylic acid and methionine residues would be exploited for a lower affinity transition site. The conserved residues Met 148 and Met 717 may serve to guide copper along the translocation pathway: from the entry/to the exit, respectively.

Met 717 of the binding site I is ~ 6 Å away. The final $2F_o - F_c$ and the unbiased $F_o - F_c$ electron-density map (before K^+ modelling) are contoured at 1.0σ (blue) and 4.0σ (green), respectively. **c**, Proposed models of Cu^+ binding at the E2/E1 transition, E1P state and the E2- P_i transition state (reported here) of the transmembrane domain of CopA. The models are based on structural alignments to SERCA1a structures, as conducted earlier for H^+ -ATPase¹².

The CopA structure brings to mind a stepwise shuttle mechanism of copper transport as observed in the CusA RND efflux transporter⁴⁷. In CopA it involves three putative stages: two transient sites (entry/exit) at the membrane interfaces, as well as the presumably high-affinity binding sites in the membrane. On the basis of the well-characterized reaction cycle of SERCA1a, these Cu^+ binding sites are expected to change conformation and modulate their apparent affinities with the Cu^+ -ATPase cycle.

Menkes' and Wilson's disease mutations

The LpCopA structure is a valuable prototype of Cu^+ -ATPases including the human proteins associated with Menkes' and Wilson's diseases (ATP7A and ATP7B, respectively). Missense mutations of these proteins are almost exclusively distributed in the conserved and now structurally determined core of the enzyme⁴⁸. A plot of known missense mutations in ATP7A (associated with the X-linked Menkes' disease) shows clusters at regions that are highly relevant for structure and function, targeting residues involved in phosphorylation/dephosphorylation, domain interactions and the ion transport pathway (Fig. 5a and Supplementary Table 2). This is even clearer when disease severity and the effects of impaired biogenesis (on subcellular localization) are taken into account (Fig. 5b). For the two mildest variants of Menkes' disease, occipital horn syndrome (OHS) and distal hereditary motor neuropathy (dHMN), the affected amino acid residues are located at the protein surface (for OHS) and at the extracellular side (for dHMN). These are positions likely to have a reduced impact on basic protein function when mutated. We have gathered genetic data from patients suffering from Menkes' disease and found 14 novel sites (incorporated in Fig. 5, listed in Supplementary Table 3) of missense mutations, including Gly728Asp of ATP7A, which is equivalent to Gly 130 of the GG kink motif in LpCopA. Our analysis pinpoints the GG kink motif as a hotspot for mutations, as the ATP7A Gly727Arg mutation (LpCopA Gly 129) is found in 4.7% of 468 independent Menkes' disease patients referred to the Kennedy Center (with a total of 383 different disease mutations observed). The immediate environment surrounding the GG kink motif, including the residues equivalent to LpCopA Arg 84, Ala 88 and Val 126, are

11. Shinoda, T., Ogawa, H., Cornelius, F. & Toyoshima, C. Crystal structure of the sodium-potassium pump at 2.4 Å resolution. *Nature* **459**, 446–450 (2009).
12. Pedersen, B. P., Buch-Pedersen, M. J., Morth, J. P., Palmgren, M. G. & Nissen, P. Crystal structure of the plasma membrane proton pump. *Nature* **450**, 1111–1114 (2007).
13. Albers, R. W. Biochemical aspects of active transport. *Annu. Rev. Biochem.* **36**, 727–756 (1967).
14. Post, R. L., Hegvary, C. & Kume, S. Activation by adenosine triphosphate in the phosphorylation kinetics of sodium and potassium ion transport adenosine triphosphatase. *J. Biol. Chem.* **247**, 6530–6540 (1972).
15. Argüello, J. M., Eren, E. & Gonzalez-Guerrero, M. The structure and function of heavy metal transport P1B-ATPases. *Biometals* **20**, 233–248 (2007).
16. Banci, L. *et al.* Affinity gradients drive copper to cellular destinations. *Nature* **465**, 645–648 (2010).
17. Gonzalez-Guerrero, M. & Argüello, J. M. Mechanism of Cu⁺-transporting ATPases: soluble Cu⁺ chaperones directly transfer Cu⁺ to transmembrane transport sites. *Proc. Natl Acad. Sci. USA* **105**, 5992–5997 (2008).
18. Gonzalez-Guerrero, M., Eren, E., Rawat, S., Stemmler, T. L. & Argüello, J. M. Structure of the two transmembrane Cu⁺ transport sites of the Cu⁺-ATPases. *J. Biol. Chem.* **283**, 29753–29759 (2008).
19. Silver, S., Nucifora, G., Chu, L. & Misra, T. K. Bacterial resistance ATPases: primary pumps for exporting toxic cations and anions. *Trends Biochem. Sci.* **14**, 76–80 (1989).
20. Argüello, J. M. Identification of ion-selectivity determinants in heavy-metal transport P1B-type ATPases. *J. Membr. Biol.* **195**, 93–108 (2003).
21. Forbes, J. R., Hsi, G. & Cox, D. W. Role of the copper-binding domain in the copper transport function of ATP7B, the P-type ATPase defective in Wilson disease. *J. Biol. Chem.* **274**, 12408–12413 (1999).
22. Morin, I., Gudin, S., Mintz, E. & Cuillel, M. Dissecting the role of the N-terminal metal-binding domains in activating the yeast copper ATPase *in vivo*. *FEBS J.* **276**, 4483–4495 (2009).
23. Williams, L. E. & Mills, R. F. P_{1B}-ATPases—an ancient family of transition metal pumps with diverse functions in plants. *Trends Plant Sci.* **10**, 491–502 (2005).
24. Solioz, M., Abicht, H. K., Mermod, M. & Mancini, S. Response of gram-positive bacteria to copper stress. *J. Biol. Inorg. Chem.* **15**, 3–14 (2010).
25. Vulpe, C., Levinson, B., Whitney, S., Packman, S. & Gitschier, J. Isolation of a candidate gene for Menkes disease and evidence that it encodes a copper-transporting ATPase. *Nature Genet.* **3**, 7–13 (1993).
26. Chelly, J. *et al.* Isolation of a candidate gene for Menkes disease that encodes a potential heavy metal binding protein. *Nature Genet.* **3**, 14–19 (1993).
27. Mercer, J. F. *et al.* Isolation of a partial candidate gene for Menkes disease by positional cloning. *Nature Genet.* **3**, 20–25 (1993).
28. Zheng, Z. *et al.* Altered microglial copper homeostasis in a mouse model of Alzheimer's disease. *J. Neurochem.* **114**, 1630–1638 (2010).
29. Leonhardt, K., Gebhardt, R., Mossner, J., Lutsenko, S. & Huster, D. Functional interactions of Cu-ATPase ATP7B with cisplatin and the role of ATP7B in the resistance of cells to the drug. *J. Biol. Chem.* **284**, 7793–7802 (2009).
30. Sazinsky, M. H., Mandal, A. K., Argüello, J. M. & Rosenzweig, A. C. Structure of the ATP binding domain from the *Archaeoglobus fulgidus* Cu⁺-ATPase. *J. Biol. Chem.* **281**, 11161–11166 (2006).
31. Sazinsky, M. H., Agarwal, S., Argüello, J. M. & Rosenzweig, A. C. Structure of the actuator domain from the *Archaeoglobus fulgidus* Cu⁺-ATPase. *Biochemistry* **45**, 9949–9955 (2006).
32. Tsuda, T. & Toyoshima, C. Nucleotide recognition by CopA, a Cu⁺-transporting P-type ATPase. *EMBO J.* **28**, 1782–1791 (2009).
33. Chintalapati, S., Al Kurdi, R., van Scheltinga, A. C. & Kuhlbrandt, W. Membrane structure of CtrA3, a copper-transporting P-type-ATPase from *Aquifex aeolicus*. *J. Mol. Biol.* **378**, 581–595 (2008).
34. Wu, C. C., Rice, W. J. & Stokes, D. L. Structure of a copper pump suggests a regulatory role for its metal-binding domain. *Structure* **16**, 976–985 (2008).
35. Kim, E. H., Charpentier, X., Torres-Urquidí, O., McEvoy, M. M. & Rensing, C. The metal efflux island of *Legionella pneumophila* is not required for survival in macrophages and amoebas. *FEMS Microbiol. Lett.* **301**, 164–170 (2009).
36. Toyoshima, C., Sasabe, H. & Stokes, D. L. Three-dimensional cryo-electron microscopy of the calcium ion pump in the sarcoplasmic reticulum membrane. *Nature* **362**, 469–471 (1993).
37. Hatori, Y., Majima, E., Tsuda, T. & Toyoshima, C. Domain organization and movements in heavy metal ion pumps: papain digestion of CopA, a Cu⁺-transporting ATPase. *J. Biol. Chem.* **282**, 25213–25221 (2007).
38. Banci, L., Bertini, I., Ciofi-Baffoni, S., Huffman, D. L. & O'Halloran, T. V. Solution structure of the yeast copper transporter domain Ccc2a in the apo and Cu(I)-loaded states. *J. Biol. Chem.* **276**, 8415–8426 (2001).
39. Jones, C. E., Daly, N. L., Cobine, P. A., Craik, D. J. & Dameron, C. T. Structure and metal binding studies of the second copper binding domain of the Menkes ATPase. *J. Struct. Biol.* **143**, 209–218 (2003).
40. Arnesano, F., Banci, L., Bertini, I., Huffman, D. L. & O'Halloran, T. V. Solution structure of the Cu(I) and apo forms of the yeast metallochaperone, Atx1. *Biochemistry* **40**, 1528–1539 (2001).
41. Boal, A. K. & Rosenzweig, A. C. Crystal structures of cisplatin bound to a human copper chaperone. *J. Am. Chem. Soc.* **131**, 14196–14197 (2009).
42. Daiho, T., Yamasaki, K., Danko, S. & Suzuki, H. Critical role of Glu40-Ser48 loop linking actuator domain and first transmembrane helix of Ca²⁺-ATPase in Ca²⁺ deocclusion and release from ADP-insensitive phosphoenzyme. *J. Biol. Chem.* **282**, 34429–34447 (2007).
43. Gonzalez-Guerrero, M., Hong, D. & Argüello, J. M. Chaperone-mediated Cu⁺ delivery to Cu⁺ transport ATPases: requirement of nucleotide binding. *J. Biol. Chem.* **284**, 20804–20811 (2009).
44. Inesi, G., Ma, H., Lewis, D. & Xu, C. Ca²⁺ occlusion and gating function of Glu309 in the ADP-fluoroaluminate analog of the Ca²⁺-ATPase phosphoenzyme intermediate. *J. Biol. Chem.* **279**, 31629–31637 (2004).
45. Heijne, G. The distribution of positively charged residues in bacterial inner membrane proteins correlates with the trans-membrane topology. *EMBO J.* **5**, 3021–3027 (1986).
46. Banci, L. *et al.* Copper(I)-mediated protein-protein interactions result from suboptimal interaction surfaces. *Biochem. J.* **422**, 37–42 (2009).
47. Long, F. *et al.* Crystal structures of the CusA efflux pump suggest methionine-mediated metal transport. *Nature* **467**, 484–488 (2010).
48. de Bie, P., Muller, P., Wijmenga, C. & Klomp, L. W. Molecular pathogenesis of Wilson and Menkes disease: correlation of mutations with molecular defects and disease phenotypes. *J. Med. Genet.* **44**, 673–688 (2007).
49. Petřek, M. *et al.* CAVER: a new tool to explore routes from protein clefts, pockets and cavities. *BMC Bioinformatics* **7**, 316 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Pauluhn, C. Schulze-Bries, T. Tomizaki and V. Olieric (SLS), T. Ursby, M. Thunnissen, J. Unge and D. Haase (MAXLAB), and U. Müller, M. Weiss and K. Paithankar (BESSY) for assistance with synchrotron data collection. Support was provided by the Dancatt program of the Danish Natural Science Research Council. We also thank C. Buchrieser for supplying the *L. pneumophila* Philadelphia genome; X. D. Su for discussions, A. M. Nielsen for technical assistance and J. L. Karlsen for support on crystallographic computing. We are thankful to T. Deva and K. Faxén for preliminary functional analysis of the protein. P.G. was supported by the Swedish Research Council, X.-Y.L. by the China Scholarship Council and J.P.M. and B.P.P. by the Carlsberg Foundation. P.N. was supported by an advanced research grant (Biomemos) of the European Research Council and at earlier stages by a Hallas-Møller stipend of the Novo Nordisk Foundation.

Author Contributions P.G. initiated the project, designed the expression construct and developed the protein production protocol assisted by J.P.M. Protein purification, activity measurements, crystallization, data collection, structure determination, refinement, and overall analysis of results were designed and performed by P.G. and X.-Y.L. jointly. B.P.P. designed and performed the Molecular Replacement screening procedure to initiate phasing, and assisted in structure determination, refinement and structural analysis. T.S. and L.B.M. identified genetic data and collected phenotypic data from Menkes' disease patients. P.N. designed and supervised the project, and analysed results. P.G., X.-Y.L., B.P.P. and P.N. wrote the paper and all authors commented on the paper.

Author Information Atomic coordinates and structure factors for the *L. pneumophila* Lpg1024 CopA crystal structure have been deposited at the Protein Data Bank under accession code 3RFU. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to L.B.M. (lbm@kennedy.dk) or P.N. (pn@mb.au.dk).

METHODS

Sample selection. Class IB P-type ATPases from *Legionella pneumophila* were selected because of their documented high expression levels in *E. coli*⁵⁰. The genome from the Philadelphia strain was provided by C. Buchrieser. Putative Cu⁺-ATPase genes *lpg0231*, *lpg1024*, *lpg1626* and *lpg2691* were cloned into pET22b(+) and checked for expression. Lpg0231, Lpg1024 and Lpg1626 were purified by nickel affinity and size-exclusion chromatography and then tested in crystallization experiments. Lpg1024 provided promising hit conditions.

Sample preparation. A pET22b(+) construct containing the full-length *lpg1024* gene was transformed into *E. coli* C43 cells. The cells were grown to an optical density at 600 nm (OD₆₀₀) of 0.6–0.8 in LB medium, induced with 1 mM IPTG, harvested after 16 h culture at 20 °C and then re-suspended in 50 mM Tris-HCl, pH 7.6, 200 mM KCl, 20% glycerol and frozen at –20 °C. Selenomethionine derivatized protein was produced from the same construct and *E. coli* strain grown in a minimal medium⁵¹, including 50 µg ml^{–1} L-SeMet. Washed cells from a pre-culture were diluted to OD₆₀₀ = 0.8, adapted to 20 °C for 1 h in a shaker and then induced with 1 mM IPTG for 16 h. Before breakage of cells with native or selenomethionine derivatized protein, 5 mM of fresh β-mercaptoethanol (BME), 1 mM phenylmethylsulphonyl fluoride, 2 µg ml^{–1} DNase I and Roche protease inhibitor cocktail (1 tablet for 8 L cells) were added to the cells. Cells were opened with an Avestin high-pressure homogenizer by three runs at 15,000–20,000 p.s.i. and kept at 4 °C throughout the purification until crystallization. Large aggregates were removed by centrifugation at 20,000g for 45 min. Membranes were isolated by centrifugation at 250,000g for 3 h, and re-suspended in 15 ml g^{–1} in 20 mM Tris-HCl, pH = 7.6, 200 mM KCl, 20% glycerol, 5 mM BME and 1 mM MgCl₂. The membrane proteins were solubilized by addition of solid octaethylene glycol monododecyl ether (C₁₂E₈) at 18.6 mM final concentration. Solubilization was performed by gentle stirring for 60 min. Unsolubilized material was removed by ultracentrifugation for 1 h at 250,000g. Solid KCl was added to solubilized membrane solution to a final concentration of 500 mM, imidazole was added to a final concentration of 50 mM and the solution was mixed with pre-equilibrated Ni²⁺ beads and incubated for at least 1 h. The beads were packed into a column and washed with 20 mM MOPS-KOH pH = 7.4, 200 mM KCl, 20% glycerol, 5 mM BME, 1 mM MgCl₂ and 0.28 mM C₁₂E₈. Bound proteins were eluted by a two-step gradient, from 0 to 250 mM and then to 500 mM imidazole. Eluted protein was checked by SDS-PAGE and the fractions containing LpCopA were pooled and concentrated to around 20–25 mg ml^{–1}. Then, typically 5 mg concentrated protein was applied to a Superose 6 size-exclusion column equilibrated in 20 mM MOPS-KOH pH = 6.8, 80 mM KCl, 20% glycerol, 5 mM BME, 1 mM MgCl₂ and 0.28 mM C₁₂E₈ (buffer A). The fractions containing LpCopA were pooled and concentrated to 20–25 mg ml^{–1}, flash frozen in aliquots of 200 µl in liquid nitrogen and stored at –80 °C. For final samples, aliquots were thawed, diluted to 10 mg ml^{–1} in buffer A and re-lipidated for 16 h with saturating amounts of dioleoyl-phosphatidylcholine lipid using an additional approximately 0.5 mg C₁₂E₈ per 100 µl LpCopA⁵². Before crystallization experiments the sample was ultracentrifuged for 10 min at 100,000g, and 10 mM NaF, 2 mM AlCl₃, 2 mM EGTA, 10 µM ammonium-tetrathiomolybdate (TTM) was added to the supernatant as well as the secondary detergents Cymal-6 or deoxy-Big-CHAP (both between 3–5× critical micelle concentration (CMC) except for selenomethionine derivatized protein that required lower concentrations). A construct without the HMBD (lacking the first 74 amino acid residues) was grown and purified in a similar manner as the full-length construct, but crystals were not obtained.

Functional characterization. The LpCopA ATPase activity was measured by the Baginsky method with Bismuth detection under aerated conditions⁵³. In a total volume of 50 µl, 15 µg (4 µM final concentration) of LpCopA was mixed with 40 mM MOPS-KOH pH = 6.8, 150 mM NaCl, 5 mM KCl, 5 mM MgCl₂, 20 mM (NH₄)₂SO₄, 1 mg ml^{–1} *E. coli* total lipid extract, 3.7 mM C₁₂E₈, 20 mM cysteine, 5 mM NaN₃ and 0.25 mM Na₂MoO₄. 5 mM ATP was added to start the reaction and the mixture was incubated at 37 °C for 15 min. 75 µl of freshly prepared stop solution (2.86% ascorbic acid, 1 M HCl, 0.48% (NH₄)₂MoO₄, 2.86% SDS) was added to stop the reaction and start colour development. After 8-min incubation on ice, 125 µl of 3.5% Bismuth citrate, 1 M HCl and 3.5% sodium citrate was added to the mixture and incubated for another 30 min at 19 °C. Absorbance was measured at 710 nm. For the HMBD-truncated construct the final protein concentration was 12 µM.

Crystallization. Crystals were grown at 19 °C using the hanging-drop vapour diffusion method with a reservoir solution containing 6% (w/v) PEG6K, 10% (v/v) glycerol, 140 mM NaCl, 3% v/v t-BuOH, 5 mM BME. 1 µl protein and 1 µl precipitant were mixed and crystals from these drops appeared within 1 week and developed to full size within 4 weeks. Optimal crystals were obtained after initial rounds of optimization using same batch aliquots. About 1 g of protein was prepared through this project and more than 2,500 crystals tested at synchrotrons (about half

of these soaked with heavy-metal compounds). Maximum size crystals (20 × 80 × 300 µm³) were mounted in Litholoops (Molecular Dimensions) and flash-cooled in liquid nitrogen. Complete native data were collected at the SLS X06SA beam line taking advantage of a PILATUS 6M detector. For the K₂Pt(CN)₄ (data collected at BESSY), Na₃IrCl₆ (MAXLAB 911-5) and para-chloromercuric benzoic acid (SLS X06SA) derivatives, concentrated stock solutions were added to the crystals at a final concentration of approximately 1 mM. For Ta₆Br₁₂ (SLS X06DA), powder and 100 mM MOPS-KOH, pH = 7.4, were added to the crystals.

Data collection and processing. Data were processed and scaled with XDS⁵⁴. The crystals belonged to space group P1 with cell parameters $a = 44.1 \text{ \AA}$, $b = 72.9 \text{ \AA}$, $c = 329.6 \text{ \AA}$, $\alpha = 89.97^\circ$, $\beta = 90.04^\circ$, $\gamma = 90.22^\circ$ with four LpCopA monomers in the unit cell related by non-crystallographic symmetry leading to a strong P2₁2₁2₁ pseudosymmetry (Supplementary Table 1 and Supplementary Fig. 12). Initial low-resolution phases were obtained by a procedure exploiting a semi-automated screening of more than 5,000 molecular replacement runs (in space group P2₁2₁2₁) using PHASER⁵⁵ with systematic combination of data sets, partial models (derived from SERCA1a structures and CopA soluble domains), and search parameters, exploiting previously described rationales⁵⁶. Initial Pt and Ir heavy-atom sites were pinpointed in anomalous difference Fourier maps using the molecular replacement phases. After this, the molecular replacement phases were discarded and the HA-site coordinates used to calculate and refine unbiased experimental phases by MIRAS in SHARP⁵⁷. The MIRAS phases were further refined and extended to 3.3 Å resolution in RESOLVE⁵⁸ using solvent flattening, histogram matching and NCS averaging. The model was built in Coot⁵⁹ and O⁶⁰ using as templates the corresponding SERCA1a structure⁶¹, the *A. fulgidus* CopA N- and P-domains structure³² and the *A. fulgidus* CopA A-domain structure³¹ to guide chain tracing, and further aided by selenomethionine positions. Model refinement was performed in phenix.refine⁶². Judged from the R factors and likelihood scores the model refinement did not improve (for example, R_{free} stayed above 40%) until the space group was changed from P2₁2₁2₁ to P1 (with one angle of the P2₁2₁2₁ unit cell shifting to 90.22°, see Supplementary Table 1). The final model yielded a crystallographic R-factor of 23.5% and a free R-factor of 26.1%. The R_{free} set for P1 incorporated the P2₁2₁2₁ pseudosymmetry to avoid bias. Molprobity⁶³ evaluation of the Ramachandran plot displayed 98.8% in allowed regions (89.6% in favoured regions) and 1.2% in disallowed regions. All figures, except Fig. 4c, were prepared using Pymol (<http://www.pymol.org>).

Patients and mutation detection. The novel missense mutations were found in a cohort of Menkes' disease patients who were referred to the Kennedy Center for molecular diagnosis. Mutations were identified from genomic DNA by PCR amplification and direct sequencing of the 23 coding exons and adjacent exon-intron boundaries, using an ABI3130XL sequencer (Applied Biosystems). The sequences of the gene-specific primers can be obtained upon request.

50. Gordon, E. et al. Effective high-throughput overproduction of membrane proteins in *Escherichia coli*. *Protein Expr. Purif.* **62**, 1–8 (2008).
51. Guerrero, S. A., Hecht, H. J., Hofmann, B., Biehl, H. & Singh, M. Production of selenomethionine-labelled proteins using simplified culture conditions and generally applicable host/vector systems. *Appl. Microbiol. Biotechnol.* **56**, 718–723 (2001).
52. Gourdon, P. HiLiDe—Systematic approach to membrane protein crystallization in lipid and detergent. *Cryst. Growth Des.* **11**, 2098–2106 (2011).
53. Cariani, L., Thomas, L., Brito, J. & del Castillo, J. R. Bismuth citrate in the quantification of inorganic phosphate and its utility in the determination of membrane-bound phosphatases. *Anal. Biochem.* **324**, 79–83 (2004).
54. Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Cryst.* **26**, 795–800 (1993).
55. Storoni, L. C., McCoy, A. J. & Read, R. J. Likelihood-enhanced fast rotation functions. *Acta Crystallogr. D* **60**, 432–438 (2004).
56. Pedersen, B. P., Morth, J. P. & Nissen, P. Structure determination using poorly diffracting membrane-protein crystals: the H⁺-ATPase and Na⁺/K⁺-ATPase case history. *Acta Crystallogr. D* **66**, 309–313 (2010).
57. Vonrhein, C., Blanc, E., Roversi, P. & Bricogne, G. Automated structure solution with autoSHARP. *Methods Mol. Biol.* **364**, 215–230 (2007).
58. Terwilliger, T. C. Maximum-likelihood density modification. *Acta Crystallogr. D* **56**, 965–972 (2000).
59. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
60. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
61. Olesen, C., Sorensen, T. L., Nielsen, R. C., Moller, J. V. & Nissen, P. Dephosphorylation of the calcium pump coupled to counterion occlusion. *Science* **306**, 2251–2255 (2004).
62. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
63. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).

Structure of the human histamine H₁ receptor complex with doxepin

Tatsuro Shimamura^{1,2,3*}, Mitsunori Shiroishi^{1,2,4*}, Simone Weyand^{1,5,6}, Hirokazu Tsujimoto^{1,2}, Graeme Winter⁶, Vsevolod Katritch⁷, Ruben Abagyan⁷, Vadim Cherezov³, Wei Liu³, Gye Won Han³, Takuya Kobayashi^{1,2}, Raymond C. Stevens³ & So Iwata^{1,2,5,6,8}

The biogenic amine histamine is an important pharmacological mediator involved in pathophysiological processes such as allergies and inflammations. Histamine H₁ receptor (H₁R) antagonists are very effective drugs alleviating the symptoms of allergic reactions. Here we show the crystal structure of the H₁R complex with doxepin, a first-generation H₁R antagonist. Doxepin sits deep in the ligand-binding pocket and directly interacts with Trp 428^{6,48}, a highly conserved key residue in G-protein-coupled-receptor activation. This well-conserved pocket with mostly hydrophobic nature contributes to the low selectivity of the first-generation compounds. The pocket is associated with an anion-binding region occupied by a phosphate ion. Docking of various second-generation H₁R antagonists reveals that the unique carboxyl group present in this class of compounds interacts with Lys 191^{5,39} and/or Lys 179^{ECL2}, both of which form part of the anion-binding region. This region is not conserved in other aminergic receptors, demonstrating how minor differences in receptors lead to pronounced selectivity differences with small molecules. Our study sheds light on the molecular basis of H₁R antagonist specificity against H₁R.

Histamine is a biogenic amine and an important mediator in various physiological and pathophysiological conditions such as arousal state, allergy and inflammation^{1–3}. Histamine exerts its effects through the activation of four distinct histamine receptors (H₁, H₂, H₃ and H₄) that belong to the G-protein-coupled-receptor (GPCR) superfamily. The H₁R, originally cloned from bovine H₁R⁴, is now known to be expressed in various human tissues including airway, intestinal and vascular smooth muscle and brain². In type I hypersensitivity allergic reactions, H₁R is activated by histamine released from mast cells, which are stimulated by various antigens⁵. Many studies have been performed to develop H₁R antagonists, known generally as anti-histamines. Many of these compounds inhibit the action of histamine on H₁R to alleviate the symptoms of allergic reactions, making H₁R one of the most validated drug targets judging from the number of drugs approved⁶. H₁R shows constitutive activity, and H₁R antagonists generally act as inverse agonists for H₁R^{7,8}. Development of H₁R antagonists has progressed through two generations. First-generation drugs such as pyrilamine and doxepin (Supplementary Fig. 1) are effective H₁R antagonists. These compounds are, however, known to show considerable side effects such as sedation, dry mouth and arrhythmia, because of penetration across the blood–brain barrier and low receptor selectivity. These H₁R antagonists can bind not only to H₁R but also to other aminergic GPCRs, monoamine transporters and cardiac ion channels. Second-generation drugs such as cetirizine and olopatadine (Supplementary Fig. 1) are less sedating and in general have fewer side effects. The improved pharmacology of the second-generation zwitterionic drugs can be attributed to a new carboxylic moiety, in combination with the protonated amine, which significantly reduces brain permeability, although residual central nervous system effects are still reported⁹. The introduction of the carboxyl moiety also

improves the H₁R selectivity of these compounds, but certain second-generation H₁R antagonists—such as terfenadine—still show cardiotoxicity because of their interaction with cardiac potassium channels^{10,11}.

A first-generation H₁R antagonist, doxepin, can cause many types of side effects due to its antagonistic effects on histamine H₂ (ref. 12), serotonin 5-HT₂, α_1 -adrenergic, and muscarinic acetylcholine receptors¹³ in addition to the inhibition of the reuptake of serotonin and noradrenaline¹⁴. Although even raw homology models of GPCRs may facilitate discovery of novel ligands^{15,16}, reliable receptor structures are essential to improve the reliability of the predictions and understand the structural basis of subtype specificity. Recently determined GPCR structures have enabled structure-based approaches to modelling ligand interactions in the binding pocket^{17–23} and are already yielding novel chemotypes predicted by virtual screening of large chemical libraries^{24,25}. Here we report the 3.1 Å resolution structure of the H₁R–T4-lysozyme fusion protein (H₁R–T4L) complex with doxepin. The crystal structure reveals the atomic details of doxepin binding and its inverse agonistic activity. The H₁R crystal structure and models of second-generation H₁R antagonists will be highly beneficial for guiding rational design of ligands that do not penetrate the blood–brain barrier while maintaining H₁ selectivity.

Overall architecture of H₁R

In the H₁R construct, T4L²⁶ was inserted into the third cytoplasmic loop (intracellular loop 3 (ICL3)) (Gln 222–Gly 404) and 19 residues were truncated from the amino-terminal region (Met 1–Lys 19) (see Methods). H₁R–T4L showed similar binding affinities for H₁R antagonists and for histamine as the wild-type H₁R expressed in yeast cells (Supplementary Table 1) and in COS-7 cells²⁷. The structure of the H₁R–T4L crystals obtained in the lipidic cubic phase (see Methods)

¹Human Receptor Crystallography Project, ERATO, Japan Science and Technology Agency, Yoshidakonoe-cho, Sakyo-ku, Kyoto 606-8501, Japan. ²Department of Cell Biology, Graduate School of Medicine, Kyoto University, Yoshidakonoe-cho, Sakyo-ku, Kyoto 606-8501, Japan. ³Department of Molecular Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA. ⁴Graduate School of Pharmaceutical Sciences, Kyushu University, 3-1-1 Maidashi, Higashi-ku, Fukuoka 812-8582, Japan. ⁵Division of Molecular Biosciences, Membrane Protein Crystallography Group, Imperial College, London SW7 2AZ, UK. ⁶Diamond Light Source, Harwell Science and Innovation Campus, Chilton, Didcot, Oxfordshire OX11 0DE, UK. ⁷Skaggs School of Pharmacy and Pharmaceutical Sciences and San Diego Supercomputer Center, University of California, San Diego, La Jolla, California 92093, USA. ⁸Systems and Structural Biology Center, RIKEN, 1-7-22 Suehiro-cho Tsurumi-ku, Yokohama 230-0045 Japan.

*These authors contributed equally to this work.

was determined in complex with the H₁R antagonist doxepin at 3.1 Å resolution (Supplementary Table 2).

H₁R is structurally most similar to the aminergic receptors (Fig. 1): β_2 -adrenergic (β_2 -AR)¹⁸, β_1 -adrenergic (β_1 -AR)¹⁹ and dopamine D3 (D3R)²³ receptors, while having larger deviations from the more phylogenetically distant rhodopsin^{17,21}, A_{2A} adenosine receptor (A_{2A}AR)²⁰ and chemokine receptor CXCR4 (ref. 22) (Supplementary Table 3). H₁R also shares common motifs with other GPCRs including D(E)RY in helix III, CWxP in helix VI and NPxxY in helix VII, as well as a disulphide bond connecting extracellular loop 2 (ECL2) with the extracellular end of helix III (Cys 100^{3,25} to Cys 180; superscripts indicate residue numbers as per the Ballesteros–Weinstein scheme²⁸) but lacks the palmitoylation site at the end of helix VIII found in many other GPCRs²⁹.

Previous GPCR structures revealed that not only the residues in the transmembrane segments but also those in the loops are critical for ligand specificity^{17–23}. ECL2 connecting helices IV and V is attached to helix III through a disulphide bond between Cys 180 in ECL2 and Cys 100^{3,25} in helix III. Seven residues (Phe 168–Val 174) before the disulphide are not included in the structure, as they did not have interpretable densities. A section of ECL2, between the disulphide bridge and the extracellular end of helix V, is particularly important because it is located at the entrance to the ligand-binding pocket. This section of ECL2 contains 7 amino acids in H₁R, as compared to 5 in β_2 -AR, 4 in D3R, and 8 in A_{2A}AR. The extra length of this ECL2 section is apparently accommodated by the increased distance between the extracellular ends of helices III and V by ~1.5 Å and ~3.1 Å when compared to β_2 -AR and D3R, respectively (Fig. 1b, c). This creates more space within the ligand-binding pocket, which can now accommodate the larger second-generation H₁R antagonists, as discussed later.

Some unique features are also observed in the transmembrane segments. A conserved Pro 161^{4,59}-induced kink in helix IV forms a tight i+3 helical turn, instead of i+4 as in β_2 -AR and D3R (Fig. 2a). This tighter turn allows accommodation of a bulky Trp side chain at position 4.56, which seems to be essential for ligand specificity of aminergic GPCRs because this position is occupied by Ser in β_2 -AR

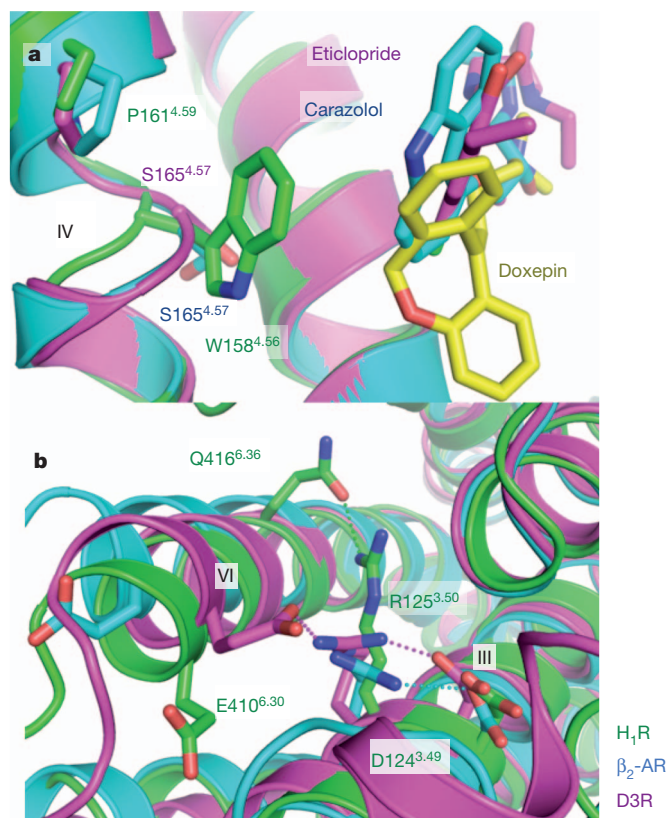


Figure 2 | Comparison of the structures of H₁R, β_2 -AR and D3R. **a**, Proline-induced kink in helix IV (H₁R, green; β_2 -AR, cyan; D3R, magenta). The side chain of Trp 158^{4,56} and Pro 161^{4,59} of H₁R and the equivalent residues of β_2 -AR (Ser 165^{4,57} and Pro 167^{4,59}) and of D3R (Ser 165^{4,57} and Pro 167^{4,59}) are also shown. **b**, Variations in the D(E)RY motif structures of H₁R, β_2 -AR and D3R are coloured in green, cyan and magenta, respectively. Side chains of Asp 124^{3,49}, Arg 125^{3,50}, Glu 410^{6,30} of H₁R and the equivalents in β_2 -AR and D3R are represented as stick models. For H₁R, Gln 416^{6,36}, which forms a hydrogen bond with Arg 125^{3,50}, is also shown. Possible hydrogen bonds are indicated by dotted lines.

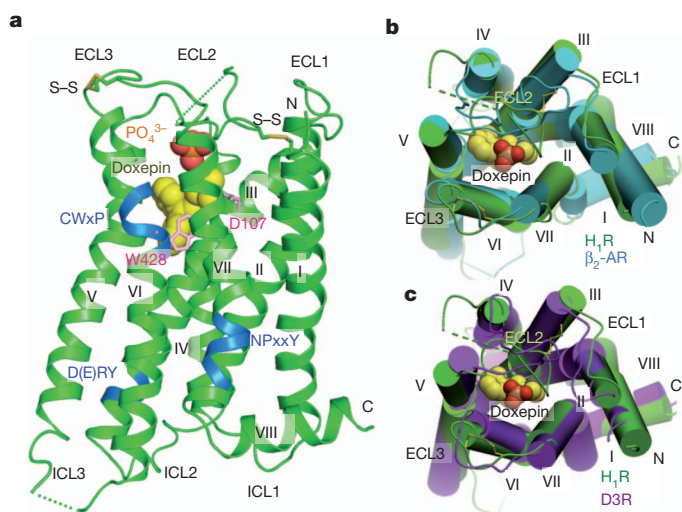


Figure 1 | Structure of H₁R complex with doxepin. **a**, Ribbon representation of the H₁R structure. Doxepin is shown as yellow spheres whereas the phosphate ion is shown as spheres with the carbon and oxygen atoms coloured orange and red, respectively. Disulphide bonds are shown as yellow sticks, and Trp 428 and Asp 107 are shown as pink sticks. Three conserved motifs D(E)RY^{3,50}, CWxP^{6,50} and NPxxY^{7,50} are highlighted in blue. **b**, Superimposition of the H₁R (green) and β_2 -AR (cyan) structures. **c**, Same as **b** but with the D3R structure coloured violet.

and D3R, and mutations of this Trp in guinea-pig H₁R to Ala, Met and Phe reduce affinity against the antagonist pyrilamine³⁰.

The ‘ionic lock’, a salt bridge between Arg^{3,50} in the conserved D(E)R^{3,50}Y motif and Asp/Glu^{6,30}, which is suggested to stabilize the inactive conformation, was observed in rhodopsin structures^{17,21} and D3R²³, but was broken in all the other GPCRs^{18–20,22}. In H₁R, Arg 125^{3,50} of the D(E)R^{3,50}Y motif does not form a salt bridge either with Glu 410^{6,30} or with Asp 124^{3,49}. Instead, the side chain of Arg 125^{3,50} adopts in a new conformer relative to previous structures, forming a hydrogen bond to Gln 416^{6,36} in helix VI (Fig. 2b). Different structures of the ionic lock regions of the receptors could be caused by modifications of ICL3. Otherwise, they might be related to the different levels of constitutive activities of the receptors.

Doxepin isomers and conformers

The doxepin used in this study contains a mixture of E and Z isomers, and each isomer can take two distinct rotational conformers of the dibenzo[b,e]oxepin ring, resulting in four distinct conformers (conformers 1–4; Supplementary Fig. 2). Two conformers, one E isomer (conformer 1) and one Z isomer (conformer 4), fit the electron density better than the other two (Supplementary Fig. 3). This result is also consistent with the *R*_{free} and the averaged *B*-factor values for each conformer (Supplementary Table 4). A 1:1 mixture of the E and Z isomers was used in the refinement. The two conformers are indistinguishable at this resolution and have nearly identical interactions

with the binding pocket; consequently, in the following sections we are referring to the E isomer, unless noted otherwise.

Ligand-binding pocket

Doxepin binds in a pocket mainly defined by the side chains of helices III, V and VI (Fig. 3a, b). Asp 107^{3,32}, a strictly conserved residue in aminergic receptors (Supplementary Table 5), forms an anchor salt bridge with the amine moiety of the ligand. This interaction has been reported to be essential for the binding of H₁R antagonists as well as agonists in mutational studies^{31–33}. This amine moiety is connected via a flexible carbon chain to the tricyclic dibenzo[b,e]oxepin ring in a hydrophobic pocket comprised of the side chains of helices III, V and VI. The tricyclic ring of doxepin sits much deeper (by ~5 Å) in the binding pocket than the ligands in the other non-rhodopsin GPCR structures (Fig. 3c). The ligand is surrounded mainly by highly conserved residues among aminergic receptors, including Ile 115^{3,40}, Phe 424^{6,44}, Trp 428^{6,48} and Phe 432^{6,52}, whereas the non-conserved residues Trp 158^{4,56} and Asn 198^{5,46} in the pocket make only minor hydrophobic interactions with doxepin (Fig. 3a, b). The importance of a large side chain at position 6.52 has been suggested for the binding of pyrilamine^{30,33}. Thr 112^{3,37} can form a hydrogen bond to the oxygen atom of the E isomer (but not the Z isomer) of doxepin, as shown in Fig. 3a, b. A suboptimal geometry and the bifurcated nature of this hydrogen bond indicate that it does not contribute significantly to binding affinity, which is supported by lack of this hydrogen-bond interaction in the H₁R–olopatadine complex, as described later. This

well-conserved pocket and its mostly hydrophobic nature should contribute to low selectivity of doxepin and other first-generation H₁R-antagonists^{13,32}. Moreover, because of its deep binding position, doxepin does not interact with ECL2, whose highly variable primary and tertiary structures are known to contribute to binding specificity of GPCR ligands³⁴.

A novel feature of the H₁R–doxepin complex is the existence of an anion-binding site at the entrance to the ligand-binding pocket (Fig. 3d). We modelled a phosphate ion, which is present at a high concentration in the crystallization buffer (300 mM ammonium phosphate), into the observed strong density in the site. This model is supported by the fact that a phosphate ion affects the binding of some ligands and the stability of H₁R (Supplementary Tables 1 and 6). The phosphate ion is coordinated by Lys 179^{ECL2}, Lys 191^{5,39}, Tyr 431^{6,51} and His 450^{7,35}; all of which, except for Tyr 431^{6,51}, are unique to H₁R (Supplementary Table 5). This encasement of the ligand in the pocket combined with an ionic interaction between the phosphate ion and the tertiary amine of doxepin (N–O distance 4.8 Å) indicates that a phosphate ion may serve as a positive modulator of ligand binding. This hypothesis has been validated by comparing thermostability (Supplementary Table 6) and ligand affinity (Supplementary Table 1) in buffers with and without phosphate. Thermostability of the receptor is increased in the presence of phosphate for all ligands except for cetirizine, which can prevent the phosphate binding, according to the modelling study discussed later. The phosphate effect is observed at a concentration as low as 1.5 mM,

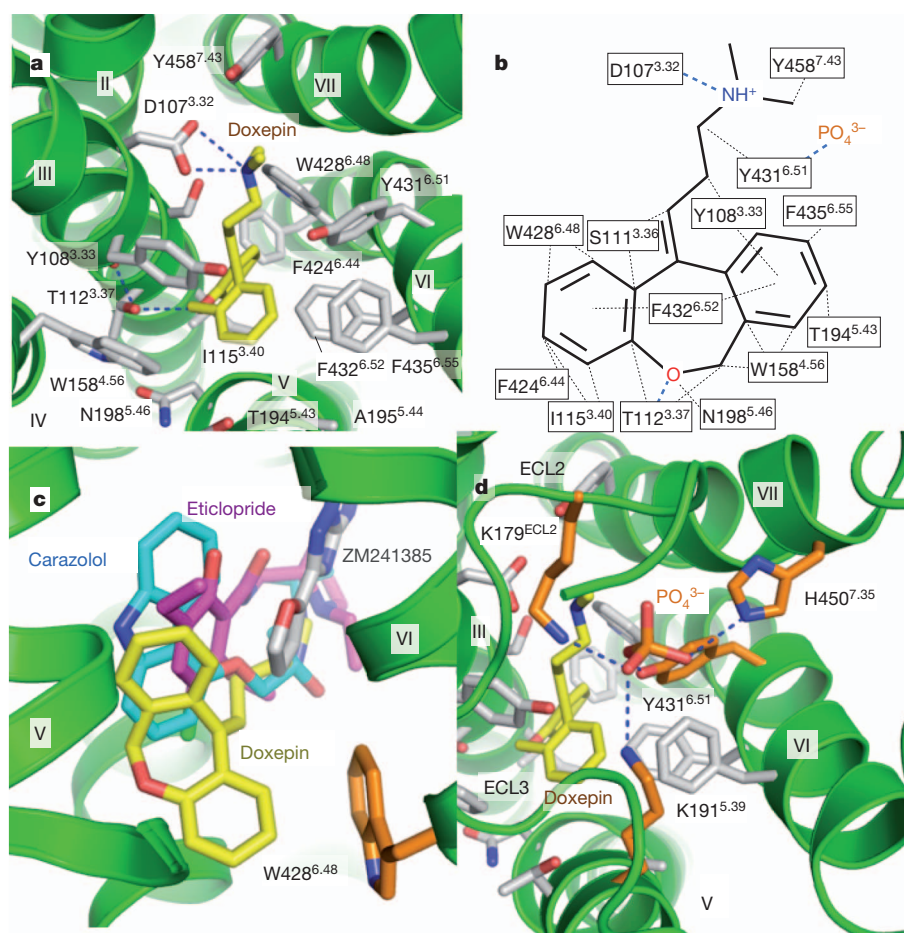


Figure 3 | Binding interactions of doxepin. **a**, Doxepin is shown as sticks with yellow carbon atoms, whereas the contact residues within 4 Å are shown with grey carbon atoms. Nitrogen and oxygen atoms are coloured blue and red, respectively. Hydrogen bonds/salt bridges are indicated as blue dotted lines. **b**, Doxepin binding interactions. Hydrophobic interactions are shown in black

dotted lines. **c**, Ligand-binding positions in non-rhodopsin GPCRs. Carbon atoms of doxepin (H₁R) are shown in yellow, carazolol (β₂-AR) in cyan, eticlopride (D3R) in magenta and ZM241385 (A_{2A}AR) in grey. **d**, Structure of the anion-binding region with a phosphate ion.

indicating its physiological relevance. The affinity of histamine and pyrilamine to the receptors also increased in the presence of phosphate.

H₁ selectivity of H₁R antagonists

Supplementary Fig. 1 lists the first and second generations of H₁R antagonists. It has been shown that second-generation H₁R antagonists are much more specific to H₁R and show much lower affinity to the other aminergic receptors^{32,35}. H₁R-antagonist specificity has been previously analysed using H₁R-homology models based on the bacteriorhodopsin or bovine rhodopsin crystal structure in combination with the H₁R antagonist pharmacophore model and mutational studies^{30,36,37}. These studies have successfully determined some residues important for selectivity, including Lys 191^{5,39}; however, contributions of the ECL residues have not been examined because these loops could not be modelled accurately based on the bacteriorhodopsin or bovine rhodopsin structure. Our H₁R structure with the extracellular loops should significantly improve understanding of H₁R-antagonist selectivity. Using flexible ligand-receptor docking^{38,39} in the ICM molecular modelling package⁴⁰ (also see Methods), we have studied H₁R selectivity for representative second-generation zwitterionic H₁R-antagonists: olopatadine, acrivastine, *R*-cetirizine (levocetirizine) and fexofenadine

(Fig. 4). Olopatadine (Fig. 4a) is a close doxepin analogue with a methyl-carboxyl substitution in one of its benzene rings. Its binding mode closely resembles doxepin, while the carboxyl group extends out of the pocket towards the extracellular space and interacts with Lys 191^{5,39} and Tyr 108^{3,33} without displacing the phosphate ion. These additional interactions can explain a reduced effect of the mutation of the conserved Asp 107^{3,32} to Ala on olopatadine binding (14-fold for olopatadine as compared to 280-fold for doxepin)^{32,41}. The orientation of the carboxyl moiety in the ECL region dictates that the oxygen atom of the dibenzo[b,e]oxepin ring is in a position where it cannot form a hydrogen bond with Thr 112^{3,37}. Although the marketed drug is only the *Z* isomer, both olopatadine *Z* and *E* isomers show similar H₁R affinities⁴¹.

Acrivastine (Fig. 4b) has a different chemical scaffold with a carboxyl group in its pyridine ring. Its longer carbon chain positions the carboxyl group higher in the ECL region, where it can form salt bridges to both Lys 191^{5,39} and Lys 179^{ECL2} amine moieties. *R*-cetirizine (Fig. 4c) has its carboxylic moiety attached directly to a piperazine amino group. The conformational modelling indicates that the carboxyl moiety can reach towards the ECL region, forming salt bridges to Lys 191^{5,39} and Lys 179^{ECL2}. Finally, fexofenadine (*R* isomer; Fig. 4d) has the most extended carboxyl-containing substituent,

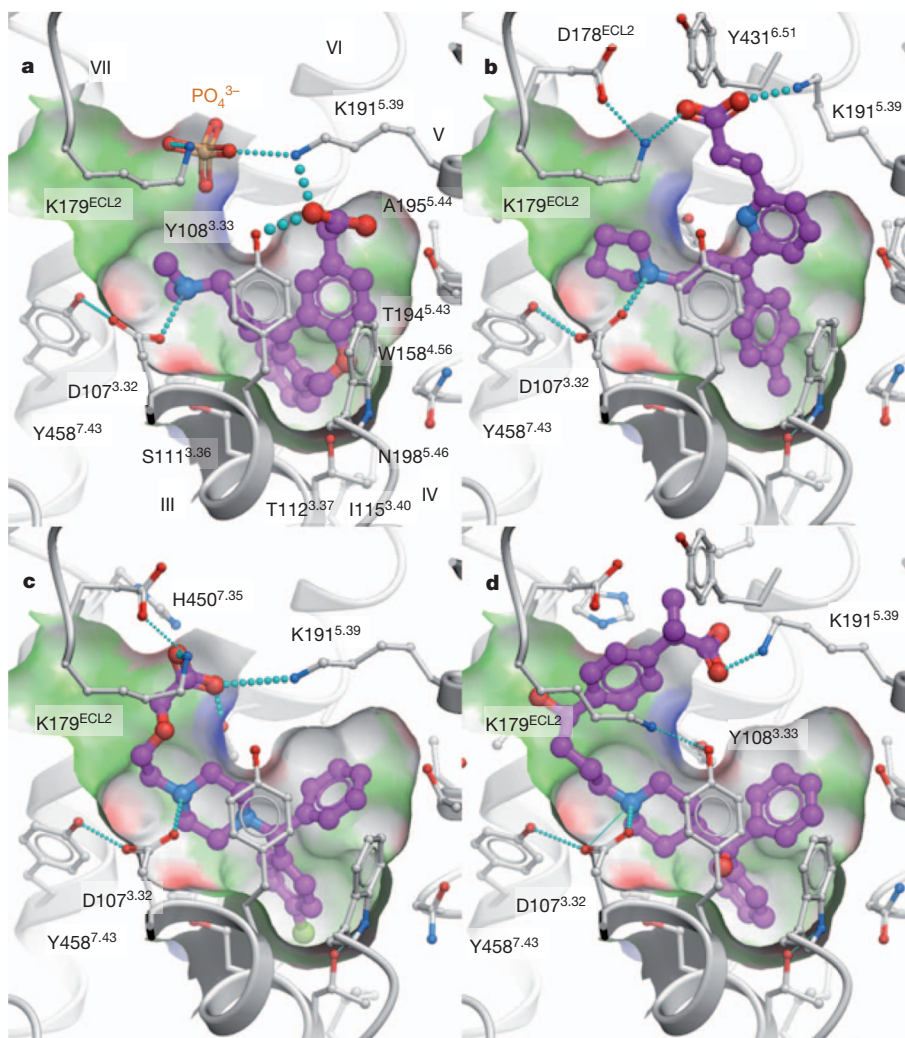


Figure 4 | Interactions of second-generation selective H₁R antagonists with the H₁R ligand-binding pocket. a–d, Conformation of each complex was predicted by global optimization of the ligand in the all-atom flexible H₁R model^{38–40} based on the H₁R–doxepin complex structure. Carbon atoms for *Z*-olopatadine co-bound with the phosphate ion (a), acrivastine

(b), *R*-cetirizine (levocetirizine) (c) and fexofenadine (d) are coloured violet. Nitrogen and oxygen atoms are coloured blue and red, respectively. Ligand contact residues of H₁R are shown with grey carbon atoms; parts of helices III, IV and ECL2 are not shown for clarity. Hydrogen bonds are shown in cyan.

which reaches outside of the binding cavity and forms a salt bridge to Lys 191^{5,39}.

Modelling of second-generation H₁R-antagonist binding to H₁R indicates that no significant protein backbone rearrangements are required to accommodate these diverse ligands. The enhanced H₁R selectivity of these compounds^{32,35} can be explained by the specific interaction of the carboxyl group with Lys residues in the ECL region, unique to H₁R. The result also shows a good agreement with earlier modelling and site-directed mutagenesis studies. Lys 191^{5,39} is known to be important for increasing affinity for some of these ligands^{30,42,43}, whereas the involvement of Lys 179^{ECL2} was suggested in the modelling study of 8R-lisuride into the ligand-binding pocket⁴⁴. Our modelling results also indicate that olopatadine is the only second-generation compound studied here for which the carboxyl moiety does not interfere with phosphate binding. The results are also supported by the fact that the presence of the phosphate ion increased the thermal stability of the H₁R–doxepin or H₁R–olopatadine complexes, whereas it did not affect the stability of the H₁R–cetirizine complex (Supplementary Table 6).

Mechanism of H₁R inactivation

H₁R antagonists act as highly effective inverse agonists of H₁R, which reduce basal activity of the receptor and therefore are expected to interfere with the key molecular switches involved in the GPCR activation mechanism. One of the switches is represented by Trp^{6,48} of the conserved CWXP^{6,50} motif, which helps to stabilize rhodopsin in its inactive dark state through a direct interaction with retinal. The recently published structure of the active-state A_{2A}AR⁴⁵ also showed that Trp^{6,48} participates in the activation-related conformational changes, where a small ligand-induced shift of Trp^{6,48} was observed in concert with the large movement of the intracellular part of helix VI. In other receptors, the role of Trp^{6,48} is less obvious; for instance, it lacks direct ligand interactions with either inverse agonists or full agonists of β_2 -AR⁴⁶. It is interesting to note that in the H₁R structure, like in inactive rhodopsin, the H₁R-antagonist doxepin does make extensive hydrophobic interactions with the Trp 428^{6,48} rings, which is unique among the known non-rhodopsin GPCR structures and could stabilize the hydrophobic packing around helix VI (Fig. 3c). Another important ligand-induced switch described in β_2 -AR is activation-related contraction of the extracellular ligand-binding pocket³⁷. Because the natural agonist histamine is much smaller than bulky H₁R antagonists, some contraction of the binding pocket is likely to accompany ligand-induced H₁R activation. Bulky compounds, capable of blocking both activation-related contraction of the pocket and the Trp 428^{6,48} switch would be very efficient in locking H₁R in an inactive conformation, which is likely to explain the as much as 78% reduction of H₁R basal activity by some H₁R antagonists⁸.

METHODS SUMMARY

H₁R–T4L was expressed in yeast *Pichia pastoris*. Ligand-binding assays were performed as described in Methods. *P. pastoris* membranes were solubilized using 1% (w/v) n-dodecyl- β -D-maltopyranoside and 0.2% (w/v) cholesteryl hemisuccinate, and purified by immobilized metal ion affinity chromatography (IMAC). After IMAC, the C-terminal GFP was cleaved by Tobacco Etch virus (TEV) protease. Then the sample mixture was passed through IMAC to remove the cleaved His-tagged GFP and TEV protease. Receptor crystallization was performed by the lipidic cubic phase (LCP) method. The protein-LCP mixture contained 40% (w/w) receptor solution, 54% (w/w) monoolein, and 6% (w/w) cholesterol. Crystals were grown in 40–50 nl protein-laden LCP boluses overlaid by 0.8 μ l of precipitant solution (26–30% (v/v) PEG400, 300 mM ammonium phosphate, 10 mM MgCl₂, 100 mM Na-citrate pH 4.5 and 1 mM doxepin) at 20 °C. Crystals were harvested directly from LCP matrix and flash frozen in liquid nitrogen. X-ray diffraction data were collected at 100 K with a beam size of 10 \times 10 μ m on the microfocus beamline I24 at the Diamond Light Source (UK). Data collection, processing, structure solution and refinement are described in Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 15 March; accepted 1 June 2011.

Published online 22 June 2011.

- Schwartz, J. C., Arrang, J. M., Garbarg, M., Pollard, H. & Ruat, M. Histaminergic transmission in the mammalian brain. *Physiol. Rev.* **71**, 1–51 (1991).
- Hill, S. J. Distribution, properties, and functional characteristics of three classes of histamine receptor. *Pharmacol. Rev.* **42**, 45–83 (1990).
- Hill, S. J. *et al.* International Union of Pharmacology. XIII. Classification of histamine receptors. *Pharmacol. Rev.* **49**, 253–278 (1997).
- Yamashita, M. *et al.* Expression cloning of a cDNA encoding the bovine histamine H₁ receptor. *Proc. Natl Acad. Sci. USA* **88**, 11515–11519 (1991).
- Simons, F. E. Advances in H₁-antihistamines. *N. Engl. J. Med.* **351**, 2203–2217 (2004).
- Overington, J. P., Al-Lazikani, B. & Hopkins, A. L. How many drug targets are there? *Nature Rev. Drug Discov.* **5**, 993–996 (2006).
- Bakker, R. A., Wieland, K., Timmerman, H. & Leurs, R. Constitutive activity of the histamine H₁ receptor reveals inverse agonism of histamine H₁ receptor antagonists. *Eur. J. Pharmacol.* **387**, R5–R7 (2000).
- Bakker, R. A., Schoonus, S. B., Smit, M. J., Timmerman, H. & Leurs, R. Histamine H₁-receptor activation of nuclear factor- κ B: roles for G β γ - and G $\alpha_{q/11}$ -subunits in constitutive and agonist-mediated signaling. *Mol. Pharmacol.* **60**, 1133–1142 (2001).
- Tashiro, M. *et al.* Dose dependency of brain histamine H₁ receptor occupancy following oral administration of cetirizine hydrochloride measured using PET with [¹¹C]doxepin. *Hum. Psychopharmacol.* **24**, 540–548 (2009).
- Woosley, R. L., Chen, Y., Freiman, J. P. & Gillis, R. A. Mechanism of the cardiotoxic actions of terfenadine. *J. Am. Med. Assoc.* **269**, 1532–1536 (1993).
- Yap, Y. G. & Camm, A. J. Potential cardiac toxicity of H₁-antihistamines. *Clin. Allergy Immunol.* **17**, 389–419 (2002).
- Okamura, N. *et al.* Functional neuroimaging of cognition impaired by a classical antihistamine, *d*-chlorpheniramine. *Br. J. Pharmacol.* **129**, 115–123 (2000).
- Cusack, B., Nelson, A. & Richelson, E. Binding of antidepressants to human brain receptors: focus on newer generation compounds. *Psychopharmacology (Berl.)* **114**, 559–565 (1994).
- Sarker, S. *et al.* The high-affinity binding site for tricyclic antidepressants resides in the outer vestibule of the serotonin transporter. *Mol. Pharmacol.* **78**, 1026–1035 (2010).
- Klabunde, T. & Hessler, G. Drug design strategies for targeting G-protein-coupled receptors. *ChemBioChem* **3**, 928–944 (2002).
- de Graaf, C., Rognan, D. & Customizing, G. Protein-coupled receptor models for structure-based virtual screening. *Curr. Pharm. Des.* **15**, 4026–4048 (2009).
- Palczewski, K. *et al.* Crystal structure of rhodopsin: a G protein-coupled receptor. *Science* **289**, 739–745 (2000).
- Cherezov, V. *et al.* High-resolution crystal structure of an engineered human β_2 -adrenergic G protein-coupled receptor. *Science* **318**, 1258–1265 (2007).
- Warne, T. *et al.* Structure of a β_1 -adrenergic G-protein-coupled receptor. *Nature* **454**, 486–491 (2008).
- Jaakola, V. P. *et al.* The 2.6 angstrom crystal structure of a human A_{2A} adenosine receptor bound to an antagonist. *Science* **322**, 1211–1217 (2008).
- Murakami, M. & Kouyama, T. Crystal structure of squid rhodopsin. *Nature* **453**, 363–367 (2008).
- Wu, B. *et al.* Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. *Science* **330**, 1066–1071 (2010).
- Chien, E. Y. *et al.* Structure of the human dopamine D₃ receptor in complex with a D₂/D₃ selective antagonist. *Science* **330**, 1091–1095 (2010).
- Kolb, P. *et al.* Structure-based discovery of β_2 -adrenergic receptor ligands. *Proc. Natl Acad. Sci. USA* **106**, 6843–6848 (2009).
- Katritch, V. *et al.* Structure-based discovery of novel chemotypes for adenosine A_{2A} receptor antagonists. *J. Med. Chem.* **53**, 1799–1809 (2010).
- Rosenbaum, D. M. *et al.* GPCR engineering yields high-resolution structural insights into β_2 -adrenergic receptor function. *Science* **318**, 1266–1273 (2007).
- Ratnala, V. R. *et al.* Large-scale overproduction, functional purification and ligand affinities of the His-tagged human histamine H₁ receptor. *Eur. J. Biochem.* **271**, 2636–2646 (2004).
- Ballesteros, J. A. & Weinstein, H. Integrated methods for the construction of three-dimensional models and computational probing of structure–function relations in G protein-coupled receptors. *Methods Neurosci.* **25**, 366–428 (1995).
- Qanbar, R. & Bouvier, M. Role of palmitoylation/depalmitoylation reactions in G-protein-coupled receptor function. *Pharmacol. Ther.* **97**, 1–33 (2003).
- Wieland, K. *et al.* Mutational analysis of the antagonist-binding site of the histamine H₁ receptor. *J. Biol. Chem.* **274**, 29994–30000 (1999).
- Ohta, K. *et al.* Site-directed mutagenesis of the histamine H₁ receptor: roles of aspartic acid¹⁰⁷, asparagine¹⁹⁸ and threonine¹⁹⁴. *Biochem. Biophys. Res. Commun.* **203**, 1096–1101 (1994).
- Nonaka, H. *et al.* Unique binding pocket for KW-4679 in the histamine H₁ receptor. *Eur. J. Pharmacol.* **345**, 111–117 (1998).
- Bruysters, M. *et al.* Mutational analysis of the histamine H₁-receptor binding pocket of histaprodifens. *Eur. J. Pharmacol.* **487**, 55–63 (2004).
- Peeters, M. C., van Westen, G. J., Li, Q. & Ijzerman, A. P. Importance of the extracellular loops in G protein-coupled receptors for ligand recognition and receptor activation. *Trends Pharmacol. Sci.* **32**, 35–42 (2011).
- Gillard, M. *et al.* H₁ antagonists: receptor affinity versus selectivity. *Inflamm. Res.* **52** (Suppl. 1), S49–S50 (2003).

36. Kiss, R., Kovari, Z. & Keseru, G. M. Homology modelling and binding site mapping of the human histamine H₁ receptor. *Eur. J. Med. Chem.* **39**, 959–967 (2004).
37. Jongejan, A. & Leurs, R. Delineation of receptor–ligand interactions at the human histamine H₁ receptor by a combined approach of site-directed mutagenesis and computational techniques—how to bind the H₁ receptor. *Arch. Pharm. (Weinheim)* **338**, 248–259 (2005).
38. Totrov, M. & Abagyan, R. Flexible protein–ligand docking by global energy optimization in internal coordinates. *Proteins* (Suppl. 1) **29**, 215–220 (1997).
39. Katritch, V. *et al.* Analysis of full and partial agonists binding to β 2-adrenergic receptor suggests a role of transmembrane helix V in agonist-specific conformational changes. *J. Mol. Recognit.* **22**, 307–318 (2009).
40. Katritch, V., Kufareva, I. & Abagyan, R. Structure based prediction of subtype-selectivity for adenosine receptor antagonists. *Neuropharmacology* **60**, 108–115 (2011).
41. Matsumoto, Y., Funahashi, J., Mori, K., Hayashi, K. & Yano, H. The noncompetitive antagonism of histamine H₁ receptors expressed in Chinese hamster ovary cells by olopatadine hydrochloride: its potency and molecular mechanism. *Pharmacology* **81**, 266–274 (2008).
42. Gillard, M., Van Der Perren, C., Moguilevsky, N., Massingham, R. & Chatelain, P. Binding characteristics of cetirizine and levocetirizine to human H₁ histamine receptors: contribution of Lys¹⁹¹ and Thr¹⁹⁴. *Mol. Pharmacol.* **61**, 391–399 (2002).
43. Leurs, R., Smit, M. J., Meeder, R., Ter Laak, A. M. & Timmerman, H. Lysine²⁰⁰ located in the fifth transmembrane domain of the histamine H₁ receptor interacts with histamine but not with all H₁ agonists. *Biochem. Biophys. Res. Commun.* **214**, 110–117 (1995).
44. Bakker, R. A. *et al.* 8R-lisuride is a potent stereospecific histamine H₁-receptor partial agonist. *Mol. Pharmacol.* **65**, 538–549 (2004).
45. Xu, F. *et al.* Structure of an agonist-bound human A_{2A} adenosine receptor. *Science* **332**, 322–327 (2011).
46. Rasmussen, S. G. *et al.* Structure of a nanobody-stabilized active state of the β ₂ adrenoceptor. *Nature* **469**, 175–180 (2011).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by the ERATO Human Receptor Crystallography Project from the Japan Science and Technology Agency and by the Targeted Proteins Research Program of MEXT (S.I.), Japan; the NIH Common Fund grant P50 GM073197 for technology development (R.C.S.) and NIH PSI:Biologics grant

U54 GM094618 (R.C.S., V.C., V.K. and R.A.); R.A. was also partly funded by NIH R01 GM071872. The work was also partly funded by the Biotechnology and Biological Sciences Research Council (BBSRC) BB/G023425/1 (S.I.), Grant-in-Aid for Challenging Exploratory Research (T.S.), the Mochida Memorial Foundation for Medical and Pharmaceutical Research (T.S. and T.K.), Takeda Scientific Foundation (M.S.) and the Sumitomo Foundation (T.K.). A part of the work was performed in the Membrane Protein Laboratory funded by the Wellcome Trust (grant 062164/Z/00/Z) at the Diamond Light Source Limited and at The Scripps Research Institute. We thank D. Axford, R. Owen and G. Evans for help with data collection at I24 of the Diamond Light Source Limited, H. Wu for help with the preparation of Supplementary Fig. 1 and Q. Xu for help on validation of data processing and A. Walker for assistance with manuscript preparation. The authors acknowledge Y. Zheng (The Ohio State University) and M. Caffrey, Trinity College (Dublin, Ireland), for the loan of the in meso robot (built with support from the NIH (GM075915), the National Science Foundation (IIS0308078), and Science Foundation Ireland (02-IN1-B266)). S.I. is thankful for the help of L. E. Johnson, a co-founder of the Diamond-MPL Project and R. Tanaka, the technical coordinator of the ERATO Human Receptor Crystallography Project.

Author Contributions T.S. purified and crystallized the receptor in LCP, optimized crystallization conditions, grew crystals for data collection, solved and refined the structure, and prepared the manuscript. M.S. designed, characterized and screened the constructs, purified the receptor, and prepared the manuscript. S.W. and S.I. collected the data and processed diffraction data with G.W.H. H.T. expressed the receptor, prepared the membrane, and performed the ligand-binding assay. V.K. and R.A. performed flexible ligand–receptor docking, and prepared the manuscript. V.C. assisted with the crystallization in LCP and prepared the manuscript. W.L. performed the thermal stability assay and assisted with the crystallization in LCP. G.W.H. refined the structure and assisted with preparing the manuscript. T.K. designed the receptor production strategy and assisted with preparing the manuscript. R.C.S. and S.I. were responsible for the overall project strategy and management and wrote the manuscript.

Author Information The coordinates and the structure factors have been deposited in the Protein Data Bank under the accession code 3RZE. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to R.C.S. (stevens@scripps.edu), S.I. (so_iwata@me.com) or T.K. (t-coba@mfour.med.kyoto-u.ac.jp).

METHODS

Construction of the H₁R expression vectors for *Pichia pastoris*. The coding sequence of the full-length human histamine H₁ receptor (H₁R-fl), in which N-linked glycosylation sites (Asn 5 and Asn 18) were mutated to glutamines, was synthesized with optimization of codon usage for *P. pastoris* (TAKARA Bio), and cloned into the pPIC9K expression vector (Invitrogen). The H₁R-T4L construct with an N-terminal 19-residue deletion and insertion of cysteine-less (C54T, C97A) T4 lysozyme into the third intracellular loop was generated by the yeast homologous recombination technique in *Saccharomyces cerevisiae* with the SmaI linearized plasmid pDDGFP2 (ref. 47) and three PCR products with ~30 bp overlapping sequences. The three fragments were individually generated by standard PCR techniques with the indicated primers. The generated plasmid integrating H₁R-T4L followed by TEV cleavage sequence (ENLYFQG), yeast enhanced GFP and octa-histidine tag (H₁R-T4L-GFP) was isolated from *S. cerevisiae*. Coding regions of the H₁R-T4L-GFP fusions were amplified by PCR using a forward primer containing a BamHI site (5'-CTAGAACTAGTGGATCCACCATG-3') and a reverse primer containing an EcoRI site (5'-GCTTGATATCGAATTCCTGCAGTTAATG-3'). The PCR products were digested with BamHI and EcoRI, and subcloned into the pPIC9K vector.

Expression and membrane preparation. The PmeI linearized pPIC9K expression vector integrating H₁R-fl-GFP or H₁R-T4L-GFP was then transformed into the *P. pastoris* SMD1163 strain by electroporation (2,000 V, 25 μ F, and 600 Ω) using a Gene Pulser I (Bio-Rad). Clone selection was performed on the YPD-agar plate containing 0.1 mg ml⁻¹ geneticin. A single colony of *P. pastoris* transformant was inoculated into BMGY medium (1% (w/v) yeast extract, 2% (w/v) peptone, 1.34% (w/v) yeast nitrogen base without amino acids, 0.00004% (w/v) biotin, 1% (w/v) glycerol, 0.1 M phosphate buffer at pH 6.0) at 30 °C with shaking at 250 r.p.m. until an OD₆₀₀ of 2–6 was reached. The cells were harvested by centrifugation. To induce expression, the cell pellet was resuspended to an OD₆₀₀ of 1.0 in BMMY medium (1% (w/v) yeast extract, 2% (w/v) peptone, 1.34% (w/v) yeast nitrogen base without amino acids, 0.00004% (w/v) biotin, 0.5% (v/v) methanol, 0.1 M phosphate buffer at pH 6.0) containing 2.5% (v/v) DMSO at 30 °C. Cells were harvested within 20 to 24 h after induction, and stored at -80 °C. Yeast cells were disrupted with 0.5 mm glass beads in a buffer containing 50 mM HEPES, pH 7.5, 120 mM NaCl, 5% (v/v) glycerol, 2 mM EDTA and EDTA-free protein inhibitor cocktail (Roche). Undisrupted cells and cell debris were separated by centrifugation at 3,000g, and yeast membranes were collected by ultracentrifugation at 100,000g for 30 min at 4 °C. Washing of the membranes was performed by repeating dounce homogenization and centrifugation in a high salt buffer containing 10 mM HEPES, pH 7.5, 1 M NaCl, 10 mM MgCl₂, 20 mM KCl and EDTA-free protease inhibitor cocktail. Prepared membranes were resuspended in a buffer containing 50 mM HEPES pH 7.5, 120 mM NaCl, 20% (v/v) glycerol and EDTA-free protease inhibitor cocktail, and snap-frozen in liquid nitrogen and stored at -80 °C until use. Membrane proteins were quantified using the bicinchoninic acid method (Pierce).

Purification of H₁R-T4L. Membrane suspension containing H₁R-T4L-GFP was thawed and incubated on ice for 30 min in the presence of 5 mM doxepin, 10 mg ml⁻¹ iodoacetamide, and EDTA-free protease inhibitor cocktail (Roche). The membrane suspension was poured into the buffer containing 20 mM HEPES pH 7.5, 500 mM NaCl, 1% (w/v) n-dodecyl-B-D-maltopyranoside (DDM, Anatrace), 0.2% (w/v) cholesteryl hemisuccinate (CHS, Sigma), 20% (v/v) glycerol and 2–3 mg ml⁻¹ membrane, and stirred gently at 4 °C for 1–2 h. The unsolubilized material was separated by centrifugation at 100,000g for 30 min. The supernatant was incubated with TALON IMAC resin (Clontech) overnight. The resin was washed with twenty column volumes of 20 mM HEPES pH 7.5, 500 mM NaCl, 10% glycerol, 0.025% (w/v) DDM, 0.005% (w/v) CHS, 100 μ M doxepin and 20 mM imidazole. The protein was eluted with 4 column volumes of 20 mM HEPES pH 7.5, 500 mM NaCl, 10% glycerol, 0.025% (w/v) DDM, 0.005% (w/v) CHS, 500 μ M doxepin and 200 mM imidazole. The eluted fractions were concentrated to 2.5 ml with a 100 kDa molecular weight cut-off AmiconUltra (Millipore). Imidazole was removed using PD-10 column (GE healthcare). Then the protein was loaded onto the Ni-Sepharose high performance resin (GE healthcare) (1.5 ml resin for ~10 mg of protein). The resin was washed with 20 column volumes of 20 mM HEPES pH 7.5, 500 mM NaCl, 10% glycerol, 0.025% (w/v) DDM, 0.005% (w/v) CHS, 500 μ M doxepin and 20 mM imidazole. The sample was eluted with 3 column volumes of 20 mM HEPES pH 7.5, 500 mM NaCl, 10% glycerol, 0.025% (w/v) DDM, 0.005% (w/v) CHS, 1 mM doxepin and 500 mM imidazole. Imidazole was removed using PD-10 column (GE healthcare). The protein was processed overnight with His-tagged TEV protease (expressed and purified in house). TEV protease and the cleaved His-tagged GFP were removed by passing the sample through the Ni-Sepharose high performance resin. The receptor was concentrated to 30–40 mg ml⁻¹ with a 100 kDa

molecular weight cut-off Vivaspinn concentrator (Vivascience). Protein purity and monodispersity were tested by SDS-PAGE and by size-exclusion chromatography using Superdex 200 (GE healthcare).

Lipidic cubic phase crystallization. Lipidic cubic phase (LCP) crystallization trials were performed using an *in meso* crystallization robot as previously described⁴⁸. Ninety-six-well glass sandwich plates were filled with 40–50 nl protein-laden LCP boluses overlaid by 0.8 μ l of precipitant solution in each well and sealed with a glass coverslip. The protein-LCP mixture contained 40% (w/w) receptor solution, 54% (w/w) monoolein, and 6% (w/w) cholesterol. Crystallization set-ups were performed at room temperature (20–22 °C). Plates were incubated and imaged at 20 °C using an automated incubator/imager (RockImager 1000, Formulatrix). Crystals were obtained in 26–30% (v/v) PEG400, 300 mM ammonium phosphate, 10 mM MgCl₂, 100 mM Na-citrate pH 4.5 and 1 mM doxepin (Sigma) (Supplementary Fig. 4). Crystals were harvested directly from LCP matrix using MiTeGen micromounts and were flash-frozen in liquid nitrogen without additional cryoprotectant.

Data collection and refinement. X-ray diffraction data were collected at 100 K with a wavelength of 0.97780 Å and with a beam size of 10 × 10 μ m on the micro-focus beamline I24 at the Diamond Light Source (UK) with a Pilatus 6M detector. Each loop was subjected to a grid scanning⁴⁹ in order to locate the crystals, which are invisible in the LCP once they are mounted. The exact locations and dimensions of the chosen crystals were determined by further grid scanning with a smaller search area. Data collection was carried out by collecting several overlapping wedges of data from adjacent positions within a single crystal. The data were processed initially with xia2⁵⁰ using Mosflm⁵¹ and Scala⁵² with the merging statistics used to determine an optimum subset of measurements to merge. The final data set consisted of data from five of the eight positions recorded, giving a total of 75 degrees of data. These data were then merged with Scala to give the final data set summarized in Supplementary Table 2. The space group was determined to be *I*422 with one molecule in the asymmetric unit. Diffraction data were slightly anisotropic, extending to 2.9 Å in the *c** direction and to 3.1 Å in the *a** and *b** directions. The structure factors up to 3.1 Å resolution were anisotropically scaled by PHASER⁵³ and then used for the subsequent molecular replacement and refinement. The structure was determined by molecular replacement with the program PHASER⁵³ using two independent search models (polyalanine of the 7 trans-membrane α -helices, and T4L) from the β_2 -AR (PDB code 2RH1) structure. We chose β_2 -AR as a model structure because it has the highest homology of trans-membrane helices with H₁R (41.7%) among the human GPCR structures. For the initial map calculation after molecular replacement, however, we used a β_2 -AR model without side chains, loops, ligand, lipids and any solvents; therefore, the final H₁R structure is not biased to the β_2 -AR structure. This is supported by low *R*_{work} and *R*_{free} values (Supplementary Table 2). All refinements were performed with REFMAC5⁵⁴ and autobUSTER⁵⁵ followed by manual examination and rebuilding of the refined coordinates in the program Coot⁵⁶. The non-lysozyme portion contains higher *B* factors (116 Å²) owing to fewer contacts as compared to T4 lysozyme (36 Å²). Calculation of the surface area buried by crystal contacts also explains this. For the non-lysozyme portion, only 8% (1,225 Å²) of 15,689 Å² solvent-accessible surface area is buried by crystal contacts. In contrast, for the T4 lysozyme portion, 32% (2,733 Å²) of the solvent accessible area (8,648 Å²) is buried by crystal interactions. Supplementary Fig. 5 also shows that there are strong interactions between T4 lysozyme domains, but relatively fewer between non-lysozyme domains throughout the crystal packing. Although the average *B* factor of the non-lysozyme domain is high as compared to T4 lysozyme, electron densities were clear for unambiguous model building (Supplementary Figs 3 and 5). The eight H₁R N-terminal residues (Thr 20–Leu 27), two C-terminal residues (Arg 486–Ser 487), and seven residues (Phe 168–Val 174) in the second extracellular loop (ECL2) are not included in the structure, as they did not have interpretable densities.

Strong and spherical electron densities (about 4 sigma) were found in the anion-binding region in the *F*_o - *F*_c omit map. We excluded the presence of a water molecule in this region owing to strong residual positive *F*_o - *F*_c densities when we modelled it as a water molecule. The coordination geometry in the highly electropositive environment surrounded by His 450^{7,35}, Lys 179^{ECL2} and Lys 191^{5,39} implied that either a phosphate or sulphate ion could be modelled. Because ammonium phosphate was added to our crystallization buffer, we modelled it as a phosphate ion. The average *B* factors of the phosphate ion and the interacting atoms are 177 Å² and 154 Å², respectively.

Ligand-binding assays. For the saturation binding experiment, yeast membrane suspensions containing H₁R-fl-GFP (20 μ g) or H₁R-T4L-GFP (5 μ g) were incubated with increasing concentrations of [³H] pyrilamine (from 0.15 to 40 nM) in a total assay volume of 200 μ l for 1 h at 25 °C. To investigate the effect of phosphate on the ligand binding, assays were performed in PBS buffer pH 7.4 (138 mM NaCl, 8.1 mM Na₂HPO₄, 27 mM KCl, 1.8 mM KH₂PO₄) or in the HEPES buffer containing 20 mM HEPES pH 7.5 and 150 mM NaCl. Nonspecific binding was

determined in the presence of 1,000-times excess unlabelled pyrilamine. Membranes were trapped on Whatman GF/B filters pre-soaked in 0.3% poly-ethylenimine, and unbound radioligands were washed with 9 ml of the PBS or HEPES buffers. The retained radioactivity was measured on an LCS-5100 liquid scintillation counter (ALOKA) in a Clearzol I scintillation liquid (Nakarai, Japan). Data were analysed by nonlinear curve-fitting with a rectangular hyperbola function using the Prism 4.0 software (GraphPad) to determine dissociation constant (K_d).

For competition binding assays, yeast membrane suspensions containing H₁R-fl-GFP or H₁R-T4L-GFP were incubated with 4 nM or 20 nM [³H]pyrilamine in the PBS buffer or the HEPES buffer in the presence of 10 nM to 100 mM histamine hydrochloride or 0.001 nM to 1 μM doxepin, or 0.01 nM to 10 μM cetirizine, pyrilamine, olopatadine and fexofenadine. Data were analysed by nonlinear curve fitting with a sigmoidal function using the Prism 4.0 to determine the half maximal inhibitory concentrations (IC₅₀). All data shown were calculated based on more than three independent experiments. Inhibition constant K_i was calculated based on the equation $K_i = IC_{50}/(1 + L/K_d)$, where L is the concentration of [³H]pyrilamine with the dissociation constant K_d .

Thermal stability assay. N-[4-(7-diethylamino-4-methyl-3-coumarinyl)phenyl]maleimide (CPM) dye was purchased from Invitrogen and dissolved in DMSO (Sigma) at 4 mg ml⁻¹ as the stock solution for future use. The stock solution was kept at -80 °C and was diluted 1:40 in dye dilution solution (10 mM buffer, 500 mM NaCl, 10% glycerol, 0.025% DDM and 0.005% CHS) before use. The thermal denaturation assay was performed with a total volume of 200 μl sample in a quartz fluorometer cuvette (Starna Cells). H₁R (4 μg) was diluted in the appropriate buffer solution to a final volume of 200 μl. Five microlitres of the diluted dye was added to the protein solution and it was incubated for 30 min at 4 °C. The mixed solution was transferred to the cuvette and the data were collected by a Cary Eclipse spectrofluorometer (Varian) with a temperature ramping rate at 1 °C min⁻¹. The excitation wavelength was 387 nm and the emission wavelength was 463 nm. All assays were performed over a temperature range starting from 20 °C to 80 °C. The stability data were processed with GraphPad Prism program (Graphpad Software). To determine the melting temperature (T_m), a Boltzmann sigmoidal equation was used to fit to the data.

Flexible ligand-receptor docking. Docking of ligands was performed using the all-atom flexible receptor docking algorithm in the ICM-Pro molecular modelling package⁵⁷ as described previously^{39,58}. The initial H₁R model was generated in ICM by building hydrogen atoms for the crystal structure of H₁R. Internal coordinate (torsion) movements were allowed in the side chains of the binding pocket, defined as residues within 8 Å distance of doxepin in the H₁R–doxepin complex. Other side chains and the backbone of the protein were kept as in the crystal structure. An initial conformation for each of the ligands was generated by Cartesian optimization of the ligand model in MMFF force field. Docking was performed by placing the ligand in a random position within 5 Å from the binding

pocket and global optimization of the complex conformational energy. The global energy of the complex was calculated as a sum of van der Waals, electrostatic, hydrogen-bonding and torsion stress terms. Stochastic global energy optimization of the complex was performed using the ICM Monte Carlo procedure with minimization⁵⁹. To facilitate side-chain rotamer switches in flexible H₁R models, the first 10⁶ steps of the Monte Carlo procedure used 'soft' van der Waals potentials and high Monte Carlo temperature, followed by another 10⁶ steps with 'exact' van der Waals method and gradually decreasing temperature. A harmonic 'distance restraint' has been applied between the amino group of the ligand and the carboxyl of the Asp 107 side chain in the initial 10⁶ steps to facilitate formation of the known salt-bridge interaction between these two groups. At least ten independent runs of the docking procedure were performed for each H₁R ligand. The docking results were considered 'consistent' when at least 80% of the individual runs resulted in conformations clustered within a root mean squared deviation of <0.5 Å to the overall best energy pose of the ligand.

47. Newstead, S., Kim, H., von Heijne, G., Iwata, S. & Drew, D. High-throughput fluorescent-based optimization of eukaryotic membrane protein overexpression and purification in *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA* **104**, 13936–13941 (2007).
48. Cherezov, V., Peddi, A., Muthusubramaniam, L., Zheng, Y. F. & Caffrey, M. A robotic system for crystallizing membrane and soluble proteins in lipidic mesophases. *Acta Crystallogr. D* **60**, 1795–1807 (2004).
49. Aishima, J. et al. High-speed crystal detection and characterization using a fast-readout detector. *Acta Crystallogr. D* **66**, 1032–1035 (2010).
50. Winter, G. xia2: an expert system for macromolecular crystallography data reduction. *J. Appl. Cryst.* **43**, 186–190 (2010).
51. Leslie, A. G. W. Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 ESF-EACMB Newslet. Protein Crystallogr.* No. 26 (1992).
52. Evans, P. Scaling and assessment of data quality. *Acta Crystallogr. D* **62**, 72–82 (2006).
53. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674 (2007).
54. Skubák, P., Murshudov, G. N. & Pannu, N. S. Direct incorporation of experimental phase information in model refinement. *Acta Crystallogr. D* **60**, 2196–2201 (2004).
55. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).
56. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and Development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
57. Abagyan, R., Orry, A., Raush, E. & Totrov, M. ICM Manual v.3.0 (MolSoft LLC, 2011).
58. Totrov, M. & Abagyan, R. Derivation of sensitive discrimination potential for virtual ligand screening. *Proceedings of the third annual international conference on computational molecular biology* 312–317 (1999).
59. Abagyan, R. & Totrov, M. Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins. *J. Mol. Biol.* **235**, 983–1002 (1994).

Deep winds beneath Saturn's upper clouds from a seasonal long-lived planetary-scale storm

A. Sánchez-Lavega¹, T. del Río-Gaztelurrutia¹, R. Hueso¹, J. M. Gómez-Forrellad², J. F. Sanz-Requena³, J. Legarreta⁴, E. García-Melendo^{2,5}, F. Colas⁶, J. Lecacheux⁷, L. N. Fletcher⁸, D. Barrado-Navascués⁹, D. Parker¹⁰ & the International Outer Planet Watch Team*

Convective storms occur regularly in Saturn's atmosphere^{1–4}. Huge storms known as Great White Spots, which are ten times larger than the regular storms, are rarer and occur about once per Saturnian year (29.5 Earth years). Current models propose that the outbreak of a Great White Spot is due to moist convection induced by water^{5,6}. However, the generation of the global disturbance and its effect on Saturn's permanent winds^{1,7} have hitherto been unconstrained⁸ by data, because there was insufficient spatial resolution and temporal sampling^{9–11} to infer the dynamics of Saturn's weather layer (the layer in the troposphere where the cloud forms). Theoretically, it has been suggested that this phenomenon is seasonally controlled^{5,9,10}. Here we report observations of a storm at northern latitudes in the peak of a weak westward jet during the beginning of northern springtime, in accord with the seasonal cycle but earlier than expected. The storm head moved faster than the jet, was active during the two-month observation period, and triggered a planetary-scale disturbance that circled Saturn but did not significantly alter the ambient zonal winds. Numerical simulations of the phenomenon show that, as on Jupiter¹², Saturn's winds extend without decay deep down into the weather layer, at least to the water-cloud base at pressures of 10–12 bar, which is much deeper than solar radiation penetrates.

Ground-based telescopes detected the first signs of the storm on 5 December 2010 at 21:05:22 Universal Time (UT), when a barely visible bright point emerged on Saturn at planetographic latitude $37.7 \pm 0.8^\circ$ N (Fig. 1). This is nearly simultaneous with the detection of electrical activity linked to the storm at 19:18 UT by the Radio Plasma Wave Science (RPWS) instrument on board the Cassini spacecraft¹³, when considering the travel time of light from Saturn to Earth (80 min) at this epoch. As in previous Great White Spot (GWS) outbreaks^{5,9,14,15}, the spot grew rapidly both in size and brightness, expanding from a length of $\sim 3,000$ km to 8,000 km in just one week. Following this expansion, the spot was centred at latitude $41.1 \pm 1.1^\circ$ N, and had a rotation period about Saturn of 10 h 41 min 43.6 s, implying a westward zonal wind speed of $-28.7 \pm 0.2 \text{ m s}^{-1}$ in the System III reference frame¹⁶. Two weeks after the outbreak, the GWS disturbance consisted of a bright compact spot (the 'storm head') followed to the east by a zonally expanding tail of bright clouds between latitudes 30° N and 45° N (Fig. 2). This tail formed a planetary-scale disturbance that encircled the planet in 55 days. The shape of the head of the storm and its drift rate remained unperturbed following its encounter with the disturbance, a behaviour consistent with the drag of the disturbance clouds by ambient zonal winds, in a similar way to that observed in the last GWS event on Saturn's equator in 1990^{9–11,15}.

Tracking individual cloud elements (sizes $\sim 1,000$ – $3,000$ km) during the observing period (5 December 2010 to 19 February 2011) allowed

us to derive their motions (Fig. 3; see also Supplementary Information). The drift rate of the head of the storm did not change in this period; it moved westwards 10 m s^{-1} faster than the westward-jet peak velocity.

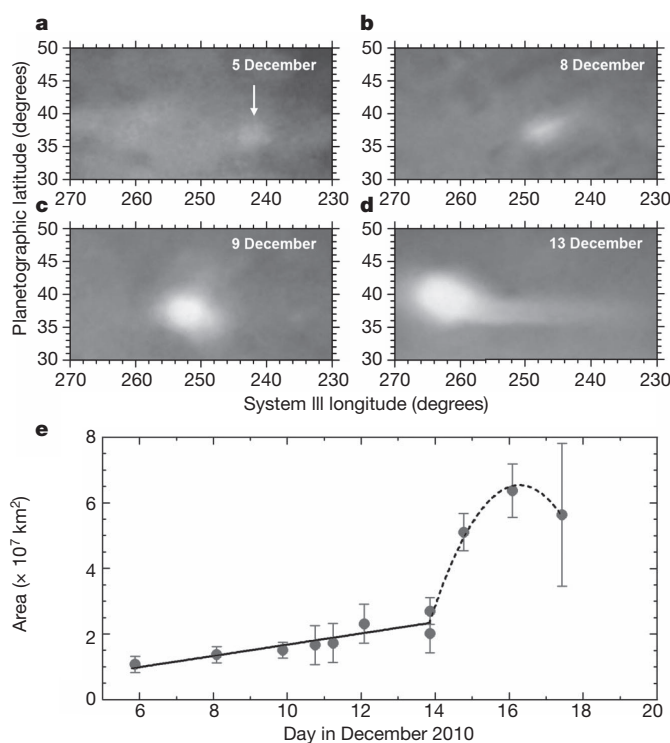


Figure 1 | Initial growth of the Great White Spot. **a–d**, Map projections of the storm clouds obtained on 5 December (**a**, by T.I.), 8 December (**b**, by S.G.), 9 December (**c**, by T.K.) and 13 December (**d**, by C.G.). Measurement of the area occupied by the bright core (without considering the southern tail in the later images) is shown in **e**. Solid and dashed lines are respectively linear and parabolic fits to the area growth rate. During the first 13 days after onset, the growth rate was $dA/dt = 20 \text{ km}^2 \text{ s}^{-1}$; during the next two days, the growth rate was $dA/dt = 212 \text{ km}^2 \text{ s}^{-1}$. The images used in this study were obtained in the visible range (wavelengths 350 nm–954 nm) by a large number of observers contributing to the IOPW-PVOL database³⁰ and the ALPO-Japan database (details are given in Supplementary Information). Multi-spectral photometric images were obtained at the 1-m telescope at Pic-du-Midi Observatory (France), the 2.2-m telescope at Calar Alto Observatory (Spain), and a 0.41-m telescope (D.P.). Error bars show the r.m.s. corresponding to several measurements of the storm area over each image.

¹Departamento de Física Aplicada I, Escuela Técnica Superior de Ingeniería, Universidad del País Vasco, Alameda Urquijo s/n, 48013 Bilbao, Spain. ²Esteve Duran Observatory Foundation, 08553 Seva, Spain. ³Universidad Europea Miguel de Cervantes, C/Padre Julio Chevalier, 47012 Valladolid, Spain. ⁴Departamento de Ingeniería de Sistemas y Automática, EUITI, Universidad País Vasco, Plaza de la Casilla 3, 48012 Bilbao, Spain. ⁵Institut de Ciències de l'Espai (CSIC-IEEC), Campus UAB, Facultat de Ciències, Torre C5, parell, 2a pl., E-08193 Bellaterra, Spain. ⁶Institut de Mécanique Céleste et de Calcul des Ephémérides, Observatoire de Paris, UMR 8028 CNRS, 77 av. Denfert-Rochereau, 75014 Paris, France. ⁷Observatoire de Paris, LESIA, 5, Place Jules Janssen, 92195 Meudon Cedex, France. ⁸Atmospheric, Oceanic and Planetary Physics, Department of Physics, Clarendon Laboratory, University of Oxford, Parks Road, Oxford OX1 3PU, UK. ⁹Observatorio de Calar Alto, Centro Astronómico Hispano Alemán, MPIA-CSIC, 04004 Almería, Spain. ¹⁰Association of Lunar and Planetary Observers (ALPO), 12911 Lerida Street, Coral Gables, Florida 33156, USA.

*Lists of participants and their affiliations appear at the end of the paper.

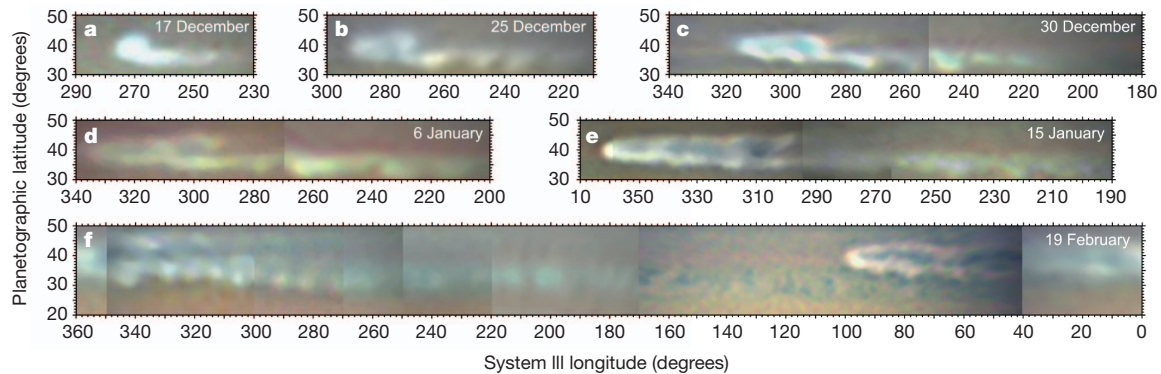


Figure 2 | Expansion of the storm clouds and the planetary-scale disturbance. Maps were made by assembling images from different observers (see Supplementary Information). The storm head moved westward (left in the maps), and showed a bow shape consistent with the meridional zonal wind profile. **a–c**, The bright clouds forming the southern branch of the disturbance (between latitudes 38° and 30° N, in a region of cyclonic vorticity) progressed eastward. Later on 22 December, a northern branch developed (latitudes 40° to 45° N, anticyclonic vorticity), which also progressed eastward (**d**, **e**). In about two months the disturbance encircled the planet, and the southern branch

On the other hand, disturbance features moved in the zonal direction with speeds very close to those of the ambient winds. Therefore winds at the upper cloud level, which have remained essentially unchanged throughout a whole Saturn year^{1,7}, were not altered by the disturbance at this early stage.

Observations obtained with a variety of filters—from the ultraviolet (375 nm) to the near-infrared (954 nm), including the deep methane absorption band at 883 nm—showed that the GWS was brighter than the adjacent undisturbed area at the same latitude by $\sim 10\%$ at 375 nm, 19% at 450 nm, 25% at 537 nm, 16% at 580 nm and 11% at 630 nm (see Supplementary Information). At 375 nm, Rayleigh scattering contributes to the brightness, but at the other wavelengths free of the gaseous methane absorption, brightness is essentially controlled by the storm particles¹⁷, suggesting that the storm injects fresh ice particles into the tropopause, where they mix with the pre-existing haze. In order to retrieve the vertical cloud structure of the storm, we used a standard

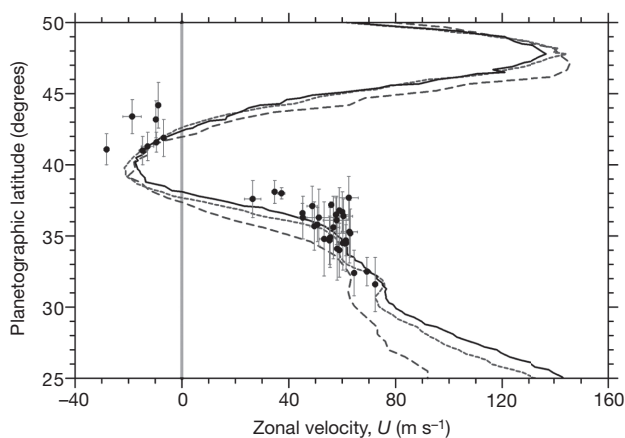


Figure 3 | Zonal winds from motions of the disturbance clouds. Data points correspond to wind speed measured for single tracers pertaining to the disturbance (Supplementary Information). The storm head corresponds to the point with the highest westward velocity (-28.7 m s^{-1}). The lines represent wind profiles; long-dashed line, obtained during the encounters of the Voyager spacecraft in 1980–81 (ref. 7); and using Cassini images at two altitude levels within the tropospheric haze (solid line) and upper cloud (short dashed) in 2004–09. The latitude uncertainty for the points is $\pm 1.3^\circ$ (s.d.); the wind speed uncertainty is $\pm 0.2 \text{ m s}^{-1}$ (s.d.) for the storm head, and ranges from ± 0.5 to $\pm 2 \text{ m s}^{-1}$ (s.d.) for other features.

elements, moving in the opposite direction, encountered the head of the storm on 29 January, 4° southward in latitude (**f**). The eastward expansion in longitude of the disturbance's central branch (between latitudes 40° and 42°) formed dark spots, one of which was persistent (probably an anticyclonic vortex) at latitude $41.9^\circ \pm 1.3^\circ$ (s.d.) with a size of $\sim 4,000 \text{ km}$ (System III longitude 308° in **e**). Small bright spots in the southern part of the disturbance at latitudes 35° to 38° N showed a periodic distribution with a dominant zonal wavelength of $15.7^\circ \pm 3^\circ$ (**b–f**) and survived for a maximum of approximately two weeks.

three-layer radiative transfer model for Saturn^{18–20}. The model atmosphere consisted of a stratospheric haze of thin Mie particles of size $\sim 0.1 \mu\text{m}$ located between 2 and 60 mbar, and a dense tropospheric haze extending from 0.1 bar down to the ammonia cloud deck at 1.4 bar. We found that the GWS was embedded within the tropospheric haze, with its cloud tops located at $\sim 150 \text{ mbar}$ —that is, 3 km below the top of the haze, and 20 km below the tropopause, which is located near 100 mbar, according to thermal infrared measurements²¹.

The rapid growth and high brightness of the GWS core, together with the abundant electric activity detected by Cassini RPWS instrument¹³, are consistent with previous proposals that these events are moist convective storms driven by condensation of ammonia and water (at the 1–2 bar and 10–12 bar levels, respectively)^{5,6}. Flow divergence at the top of the ascending motions can be related to the growing area A of the storm (Fig. 1) through mass continuity by $\nabla \cdot \mathbf{V} = (1/A)(dA/dt) = (1-5) \times 10^{-6} \text{ s}^{-1}$ (where \mathbf{V} is the horizontal velocity). This implies a vertical velocity (w) at the cloud top of the storm of $w \leq h \cdot \nabla \cdot \mathbf{V} \approx 2.5 \text{ m s}^{-1}$, where $h = 260 \text{ km}$ is the vertical length of the path followed by the ascending convective parcels from altitude (pressure P) levels $P_{\text{bottom}} \approx 10 \text{ bar}$ (water clouds) to $P_{\text{top}} \approx 0.1 \text{ bar}$ (the tropopause). However, deep moist convection in Saturn is vigorous in the core of the ascending convective towers^{5,6}, and to first order the maximum vertical velocity w_{max} can be related to the convective available potential energy (CAPE)

by $\text{CAPE} = \int_{z_1}^{z_2} g \frac{\Delta T}{T(z)} dz = \frac{w_{\text{max}}^2}{2}$, where g is the acceleration due to

gravity and ΔT is the temperature difference between the ascending parcel and the ambient (which has vertical temperature profile $T(z)$). Assuming a constant ΔT of just 1 K, this results in $w_{\text{max}} = \sqrt{2\text{CAPE}} \approx \left(2g \frac{\Delta T}{\langle T \rangle} h\right)^{1/2} \approx 150 \text{ m s}^{-1}$ for $g = 10 \text{ m s}^{-2}$ and mean temperature $\langle T \rangle = 250 \text{ K}$. Given the high vertical velocities, the large vertical extent of the core of the GWS, and its enormous horizontal size (of the order of the width of the mid-latitude jets), this huge disturbance serves as a tool to diagnose the dynamics of Saturn's weather layer, which ranges from the deep troposphere (10 bar) to 0.01 bar or higher in the stratosphere.

In order to explain the structure of the disturbance at cloud level, we ran nonlinear simulations of the response of Saturn's upper troposphere and stratosphere to a steady vertical heat source that tries to mimic the GWS convective storm head using the EPIC code²². We tested the structure of the atmosphere by varying the zonal wind

vertical profile and the thermal structure in a forward modelling of the potential vorticity (PV) field²³. As both the PV and the cloud field are considered to act as passive tracers of the flow, we sought the PV field evolution that best reproduced the observed cloud morphology. Figure 4 shows results from a simulation of the first 12 days of the PV field that closely resembles the evolution of the cloud disturbance (compare to Fig. 2a–c in this work, and to figure 3 in ref. 13). These successful simulations require a weather layer with a very low static stability, $N^2 = 0.03 \times 10^{-4} \text{ s}^{-2}$ (N is the Brunt-Väisälä frequency), which is close to a neutral profile, and winds that increase slightly with depth across it (from 0.5 to 10 bar). In addition, the wind structure in our simulation changed very little as the disturbance progressed, in agreement with what we observe at cloud level.

GWSs have been observed to occur once per Saturnian year in the northern hemisphere summertime season^{5,9,10} (see Supplementary Information). The 2010 event occurred in the much earlier springtime season within the same westward jet as the 1903 event^{5,9–11,14}, pointing to seasonal insolation as their triggering mechanism. This is puzzling, as seasonal temperature changes occur only above ~ 500 mbar altitude (as shown by solar radiation penetration in Saturn's atmosphere²⁴ and by thermal infrared measurements^{21,25}); this is much higher than the

10–12 bar level of the storm source^{5,6}. The zonal wind profile at the 0.5–1 bar level does not change over one Saturnian year^{1,7} (Fig. 3), ruling out a related seasonal dynamical instability. So the deep trigger could then be induced by seasonal changes in the upper troposphere temperature gradients directly affecting the deep vertical motions or indirectly modifying the pattern of meridional cell circulations between opposed jets proposed to exist at different altitudes in the weather layer^{1,26}.

The observed properties of the GWS and its modelling suggest that the storm is fuelled by a reservoir of water vapour at the 10–12 bar level in the westward jet, together with a sustained strong convergence^{5,6}. The storm head extends downwards from ~ 260 km below the cloud tops, and across the weather layer (0.5–10 bar), where zonal winds stay constant (or increase slightly) with depth, in agreement with both previous observations and models of mid-latitude Saturnian vortices^{27,28} and wind speed measurements at the 2–3 bar level²⁹. The huge planetary-scale disturbance GWS triggered by the storm did not modify the westward zonal jet structure substantially; this indicates that Saturn's tropospheric winds are robust and extend well below the sunlight penetration level. Taken together, our data favour the hypothesis that these tropospheric winds have a deep origin¹, as is the case for Jupiter's winds¹².

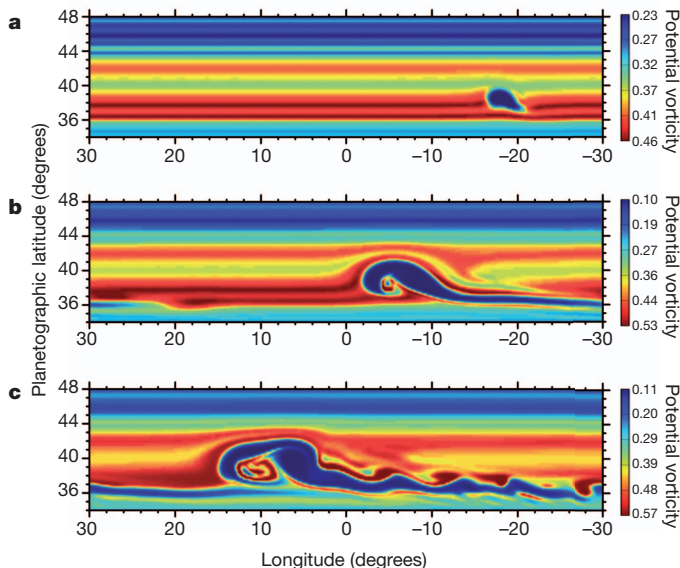


Figure 4 | Models of the GWS planetary-scale disturbance. Shown are nonlinear EPIC simulations²² of the onset and initial evolution of the potential vorticity field associated with the storm at pressure level 850 mbar at days 1 (a), 6 (b) and 12 (c). To simulate the GWS, we introduced on Saturn's wind flow a vertically extended continuous heat source column with a power of 500 W m^{-2} and a Gaussian shape of size 0.6° . The constant injection of heat from altitude levels 0.5 to 10 bar mimics the continuous latent heat released upward by the convective storm. The model atmosphere depends on the vertical and meridional structure of the zonal wind velocity $U(y, z)$, and on the vertical thermal structure represented here by the static stability or Brunt-Väisälä frequency $N(P)$, where we use z or pressure P as vertical coordinates and y as meridional coordinate. The wind field structure is modelled as a product of two functions: $U(y, p) = u_h(y)u_v(P)$, where $u_h(y)$ is the wind profile at the cloud tops ($P_0 = 500$ mbar, refs 1, 12, Fig. 3), and $u_v(P) = 1 + m \ln(P/P_0)$ is a non-dimensional vertical amplitude factor. We tested different values for slope m above and below $P_0 = 500$ mbar (refs 27, 28). Similarly, for the thermal structure we adopted $N(P)$ increasing with altitude for $P < 500$ mbar (ref. 26) and $N(P) = \text{constant}$ for $P > 500$ mbar. The horizontal domain for the simulations was a $60^\circ \times 15^\circ$ channel with resolution 0.23° per pixel, and the vertical domain consisted of 8 layers from 10 mbar to 10 bar. The simulation that best resembles the evolution of the observed cloud field was obtained for a low static stability $N^2 = 0.03 \times 10^{-4} \text{ s}^{-2}$ ($P > 500$ mbar) and for zonal winds that have slightly positive vertical shear ($m = 0.1$), that is, winds increasing below the cloud tops across the altitude range from $P = 0.5$ to 10 bar. Longitude is given in an arbitrary longitude system centred in the simulation domain.

Received 26 February; accepted 10 May 2011.

1. Del Genio, A. D. *et al.* in *Saturn from Cassini-Huygens* (eds Dougherty, M., Esposito, L. & Krimigis, T.) 113–159 (Springer, 2009).
2. Porco, C. C. *et al.* Cassini imaging science: initial results on Saturn's atmosphere. *Science* **307**, 1243–1247 (2005).
3. Dyudina, U. A. *et al.* Lightning storms on Saturn observed by Cassini ISS and RPWS during 2004–2006. *Icarus* **190**, 545–555 (2007).
4. Fischer, G. *et al.* Analysis of a giant lightning storm on Saturn. *Icarus* **190**, 528–544 (2007).
5. Sánchez-Lavega, A. & Battaner, E. The nature of Saturn's atmospheric Great White Spots. *Astron. Astrophys.* **185**, 315–326 (1987).
6. Hueso, R. & Sánchez-Lavega, A. A. Three-dimensional model of moist convection for the giant planets II: Saturn's water and ammonia moist convective storms. *Icarus* **172**, 255–271 (2004).
7. Sánchez-Lavega, A., Rojas, J. F. & Sada, P. V. Saturn's zonal winds at cloud level. *Icarus* **147**, 405–420 (2000).
8. Sayanagi, K. M. & Showman, A. P. Effects of a large convective storm on Saturn's equatorial jet. *Icarus* **187**, 520–539 (2007).
9. Sánchez-Lavega, A. *et al.* The Great White Spot and disturbances in Saturn's equatorial atmosphere during 1990. *Nature* **353**, 397–401 (1991).
10. Sánchez-Lavega, A., Lecacheux, J., Colas, F. & Laques, P. Temporal behavior of cloud morphologies and motions in Saturn's atmosphere. *J. Geophys. Res.* **98** (E10), 18857–18872 (1993).
11. Barnett, C. D., Westphal, J. A., Beebe, R. F. & Huber, L. F. Hubble Space Telescope observations of the 1990 equatorial disturbance on Saturn: zonal winds and central meridian albedos. *Icarus* **100**, 499–511 (1992).
12. Sánchez-Lavega, A. *et al.* Depth of a strong jovian jet from a planetary-scale disturbance driven by storms. *Nature* **451**, 437–440 (2008).
13. Fischer, G. *et al.* A giant thunderstorm on Saturn. *Nature* doi:10.1038/nature10205 (this issue).
14. Sánchez-Lavega, A. Motions in Saturn's atmosphere: observations before Voyager encounters. *Icarus* **49**, 1–16 (1982).
15. Beebe, R. F., Barnett, C., Sada, P. V. & Murrell, A. S. The onset and growth of the 1990 equatorial disturbance on Saturn. *Icarus* **95**, 163–172 (1992).
16. Seidelmann, P. K. *et al.* Report of the IAU/IAG working group on cartographic coordinates and rotational elements: 2006. *Celestial Mech. Dyn. Astron.* **98**, 155–180 (2007).
17. Sánchez-Lavega, A. *et al.* Large-scale storms in Saturn's atmosphere during 1994. *Science* **271**, 631–634 (1996).
18. Acarreta, J. R. & Sánchez-Lavega, A. Vertical cloud structure in Saturn's 1990 equatorial storm. *Icarus* **137**, 24–33 (1999).
19. Pérez-Hoyos, S., Sánchez-Lavega, A., French, R. G. & Rojas, J. F. Saturn's cloud structure and temporal evolution from ten years of Hubble Space Telescope Images (1994–2003). *Icarus* **176**, 155–174 (2005).
20. West, R. A., Baines, K. H., Karkoschka, E. & Sánchez-Lavega, A. in *Saturn from Cassini-Huygens* (eds Dougherty, M., Esposito, L. & Krimigis, T.) 161–179 (Springer, 2009).
21. Fletcher, L. N. *et al.* Seasonal change on Saturn from Cassini CIRS observations, 2004–2009. *Icarus* **208**, 337–352 (2010).
22. Dowling, T. E. *et al.* The explicit planetary isentropic-coordinate (EPIC) atmospheric model. *Icarus* **132**, 221–238 (1998).
23. Read, P. L. *et al.* Mapping potential vorticity dynamics on Saturn: zonal mean circulation from Cassini and Voyager data. *Planet. Space Sci.* **57**, 1682–1698 (2009).

24. Pérez-Hoyos, S. & Sánchez-Lavega, A. Solar flux in Saturn's atmosphere: maximum penetration and heating rates in the aerosol and cloud layers. *Icarus* **180**, 368–378 (2006).
25. Li, L. *et al.* Saturn's emitted power. *J. Geophys. Res.* **115**, E11002, doi:10.1029/2010JE003631 (2010).
26. Barnet, C. D., Beebe, R. F. & Conrath, B. J. A seasonal radiative-dynamic model of Saturn's troposphere. *Icarus* **98**, 94–107 (1992).
27. García-Melendo, E., Sánchez-Lavega, A. & Hueso, R. Numerical models of Saturn's long-lived anticyclones. *Icarus* **191**, 665–677 (2007).
28. del Río-Gaztelurrutia, T., Legarreta, J., Hueso, R., Pérez-Hoyos, S. & Sánchez-Lavega, A. A long-lived cyclone in Saturn's atmosphere: observations and models. *Icarus* **209**, 665–681 (2010).
29. Choi, D. S., Showman, A. P. & Brown, R. H. Cloud features and zonal wind measurements of Saturn's atmosphere as observed by Cassini/VIMS. *J. Geophys. Res.* **114**, E04007, doi:10.1029/2008JE003254 (2009).
30. Hueso, R. *et al.* The International Outer Planets Watch atmospheres node database of giant planets images. *Planet. Space Sci.* **58**, 1152–1159 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements A.S.-L., T.d.R.-G., R.H., J.L., J.M.G.-F., E.G.-M. and J.F.S.-R. are supported by the Spanish MICIIN, by FEDER and by Gobierno Vasco. We thank S. Pérez-Hoyos for initial support of this study and M. Alises and A. Guijarro for taking the Calar Alto Observatory images (CAHA and MPG/CSIC). E.G.-M. used computing facilities at CESCA (Barcelona) supported by MICIIN. L.N.F. is supported by a Glasstone fellowship at the University of Oxford. The International Outer Planet Watch (IOPW) Team and other individual contributors listed in Supplementary Information provided most of the images used for tracking in this study; these images were complemented in some cases with images taken from contributors to the ALPO Japan (Association of Lunar and Planetary Observers) database (<http://alpo-j.asahikawa-med.ac.jp/indexE.htm>).

Author Contributions A.S.-L. coordinated the study and performed the motion and wind measurements and the interpretation. T.d.R.-G. performed the photometric measurements and filter calibrations. R.H. measured the storm growth rate and with J. Legarreta coordinated the IOPW database. J.M.G.-F. prepared the map projections and image search. J.F.S.-R. performed the radiative transfer calculations. E.G.-M. and J. Legarreta performed the EPIC simulations. F.C. and J. Lecacheux provided the Pic-du-Midi photometric images. L.N.F. provided data on the thermal structure of the storm. D.B.-N. provided the photometric images obtained at Calar Alto Observatory,

and D.P. provided photometric images at selected wavelengths. All these authors discussed the results and commented on the manuscript. Contributors to the IOPW-PVOL database are listed at the end of this Letter.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.S.-L. (agustin.sanchez@ehu.es).

The International Outer Planet Watch (IOPW) Team

T. Akutsu¹, T. Barry², J. Beltran³, S. Buda⁴, B. Combs⁵, F. Carvalho⁶, P. Casquinha⁷, M. Delcroix⁸, S. Ghomizadeh¹, C. Go⁹, J. Hotershall¹⁰, T. Ikemura¹, G. Jolly¹¹, A. Kazemoto¹, T. Kumamori¹, M. Lecompte¹², P. Maxson¹³, F. J. Melillo¹⁴, D. P. Milika¹⁵, E. Morales¹⁶, D. Peach¹⁷, J. Phillips¹⁸, J. J. Poupeau¹⁹, J. Sussenbach²⁰, G. Walker²¹, S. Walker²², T. Tranter²³, A. Wesley²⁴, T. Wilson²⁵ & K. Yunoki¹

¹ALPO-Japan (Kansai branch) 520-0242, 4-8-11, Honkatata Otsu-city, Shiga, Japan, and IOPW at Escuela Técnica Superior de Ingeniería de Bilbao, Ald. Urquijo s/n 48013 Bilbao, Spain. ²406 Bromide Street, Broken Hill, New South Wales 2880, Australia. ³Dr Jose Prades n.7, 12511 Rossell, Castellon, Spain. ⁴31 Raglan Street, St Kilda East, 3183 Melbourne, Australia. ⁵606 Frank Powell Road, Buena Vista, Georgia 31803, USA ⁶Centro de Estudos do Universo, 17380-000 Brotas, Brazil. ⁷Rua Gil Eanes lte. 216, 2° dto, Portais da Arrábida, Quinta do Anjo, 2850-745 Palmela, Portugal. ⁸2, rue de l'Ardache, 31170 Tournefeuille, France. ⁹Physics Department, University of San Carlos, Nasipit, Talamban, 6000 Cebu City, Philippines. ¹⁰21 Tarcutta Street Cleveland, Brisbane, Queensland 4163, Australia. ¹¹848 S. Rosemont Ct., Gilbert, Arizona 85296, USA. ¹²51, Frogmore Close, Slough, SL1 9BN, UK. ¹³17109 N. Eureka Trail, Surprise, Arizona 85374, USA. ¹⁴14 Glen-Hollow Dr., E-16, Holtsville, New York 11742, USA. ¹⁵12 Capri Crescent, Sellicks Beach, South Australia 5174, Australia. ¹⁶HC2 PO Box 20031, Aguadilla, 00603 Puerto Rico. ¹⁷28 North Road, Selsey, West Sussex, PO20 0BW, UK. ¹⁸1280 Johnnie Dodds Blvd, Suite 104, Mt Pleasant, South Carolina 29464, USA. ¹⁹6 Grande Rue, 91470 Pecqueuse, France. ²⁰Meekrap-oord 3, 3991 VE Houten, The Netherlands. ²¹632 Byars Drive, Macon, Georgia 31210, USA. ²²294 Myrtle Street, Manchester, New Hampshire 03104, USA. ²³8 Ellalong Street, Kurri Kurri, New South Wales 2327, Australia. ²⁴PO Box 409, Campbell, Australian Capital Territory 2612, Australia. ²⁵755 North Point Parkway, Suite 219m, Alpharetta, Georgia 30022, USA.

A giant thunderstorm on Saturn

G. Fischer¹, W. S. Kurth², D. A. Gurnett², P. Zarka³, U. A. Dyudina⁴, A. P. Ingersoll⁴, S. P. Ewald⁴, C. C. Porco⁵, A. Wesley⁶, C. Go⁷ & M. Delcroix⁸

Lightning discharges in Saturn's atmosphere emit radio waves¹ with intensities about 10,000 times stronger than those of their terrestrial counterparts². These radio waves are the characteristic features of lightning from thunderstorms on Saturn, which last for days to months². Convective storms about 2,000 kilometres in size have been observed in recent years at planetocentric latitude 35° south^{3–5} (corresponding to a planetographic latitude of 41° south). Here we report observations of a giant thunderstorm at planetocentric latitude 35° north that reached a latitudinal extension of 10,000 kilometres—comparable in size to a 'Great White Spot'^{6,7}—about three weeks after it started in early December 2010. The visible plume consists of high-altitude clouds that overshoot the outermost ammonia cloud layer owing to strong vertical convection, as is typical for thunderstorms. The flash rates of this storm are about an order of magnitude higher than previous ones, and peak rates larger than ten per second were recorded. This main storm developed an elongated eastward tail with additional but weaker storm cells that wrapped around the whole planet by February 2011. Unlike storms on Earth, the total power of this storm is comparable to Saturn's total emitted power. The appearance of such storms in the northern hemisphere could be related to the change of seasons⁷, given that Saturn experienced vernal equinox in August 2009.

On 5 December 2010, the Radio and Plasma Wave Science (RPWS) instrument⁸ on board the Saturn-orbiting Cassini spacecraft detected radio emissions associated with a new lightning storm. On the same day, the Cassini Imaging Science Subsystem (ISS)⁹ observed a bright cloud, 1,300 km × 2,500 km in size, at planetocentric latitude 32° N and longitude 245° W (here we use the Voyager Saturn Longitude System, SLS, corresponding to a rotation period of 10 h 39 min 22 s). The storm might have started somewhat earlier, as there is a gap in observations of about 2 days before the detection in the RPWS data.

The radio waves emitted by lightning discharges in Saturn's atmosphere are known as Saturn electrostatic discharges¹, SEDs, and are usually observed as short individual radio bursts, but in this event (see the time–frequency spectrogram of Fig. 1) they cluster to an almost continuous emission owing to the high flash rate. On their way to Cassini's radio receiver the SEDs must pass through Saturn's ionosphere, and therefore the low-frequency extent of the emission can be used as a measure of the peak electron plasma frequency of the intervening plasma¹⁰. The decreasing low-frequency cut-off in Fig. 1 is consistent with a storm rotating with the planet from the dayside to the nightside, where electron densities of the ionosphere are lower. Such measurements by the Voyager and Cassini spacecraft have yielded typical peak electron densities somewhat larger than 10⁵ cm^{−3} around noon, but nightside densities as low as 10²–10³ cm^{−3} (derived from the Voyager SEDs¹⁰) have not so far been observed by Cassini¹¹. Averaged day/night variations in electron density observed by Cassini were less than two orders of magnitude¹¹. Models taking into account an influx of water from Saturn's rings

can reproduce the electron densities at dusk and dawn, but most models cannot reproduce the large day/night variation; this is because of the long-lived H⁺ ion in Saturn's ionosphere^{11–13}.

Viewed from Cassini, SEDs occur in episodes, and start or stop at the time when the visible storm enters or leaves the radio horizon. This consistency in longitudes enables an association of SEDs with storm clouds. Further confirmation of such a link comes from a drift rate of the storm cloud consistent with the SED episode recurrence period, and a correlation between cloud brightness and SED rates^{4,5}. Figure 2 shows the episodic SED activity, which was comparable to previous SED storms in the first days of the storm and reached unprecedented levels on 12 December. The real flash rate is unknown, as we do not know how many SEDs would appear very close to the source. The lower limit of the peak flash rate is ~10 SEDs per second.

The polarization of SEDs for this storm at frequencies below 2 MHz is strongly left-hand circularly polarized, which is opposite to the SED polarization from southern hemisphere storms¹⁴. This difference is related to the opposite direction of the magnetic field in the two hemispheres relative to the radio wave propagation vector, as the field is

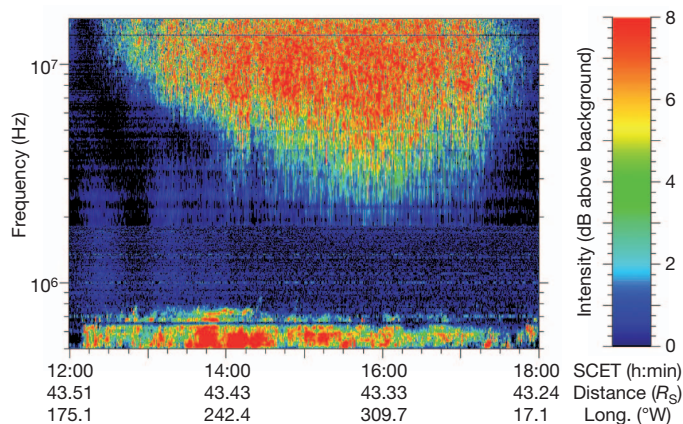


Figure 1 | Time-frequency spectrogram of the SED episode on 12 December 2010. The colour-coded intensity (with 30% background division) of the radio emissions is plotted as a function of spacecraft event time (SCET) over 6 h and frequency from 500 kHz to 16 MHz on a logarithmic scale. Cassini coordinates (distance to Saturn's centre in units of Saturn's radius, R_S , and SLS west longitude in degrees, 'Long.') are indicated on the abscissa. Cassini was in the equatorial plane at a local time of ~18.6 h. The RPWS instrument sweeps in frequency, and it detects the broadband SEDs at whatever frequency (above the ionospheric cut-off) it happens to be tuned to at the time of the flash. This SED episode shows such a high flash rate that the receiver sweep rate of ~28 frequency channels per second (35.2 ms per channel) can no longer resolve the single SEDs. Flash rates of 5–10 SEDs per second can lead to a temporal superposition of SEDs that normally extend over several frequency channels. At the edges of the episode, where the rate is lower, one can see the individual SED bursts. The continuous emission below 800 kHz is Saturn kilometric radiation.

¹Space Research Institute, Austrian Academy of Sciences, Schmiedlstrasse 6, A-8042 Graz, Austria. ²Department of Physics and Astronomy, The University of Iowa, 203 Van Allen Hall, Iowa City, Iowa 52242, USA. ³Observatoire de Paris-Meudon, 5 Place Jules Janssen, 92195 Meudon Cedex, France. ⁴Geological and Planetary Sciences, 150-21, California Institute of Technology, Pasadena, California 91125, USA. ⁵Cassini Imaging Central Laboratory for Operations, Space Science Institute, 4750 Walnut Street, Boulder, Colorado 80301, USA. ⁶82 Merryville Drive, Murrumbateman, 2582 New South Wales, Australia. ⁷Physics Department, University of San Carlos, Nasipit, Talamban, 6000 Cebu City, Philippines. ⁸Commission des observations planétaires, Société Astronomique de France, 2 rue de l'Ardèche, 31170 Tournefeuille, France.

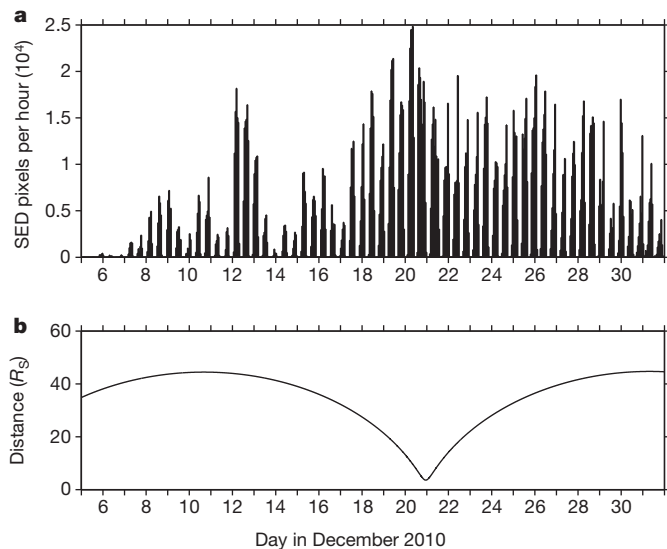


Figure 2 | Saturn's lightning activity and Cassini's distance to Saturn as a function of time for December 2010. **a**, The number of SED pixels as a function of day in December 2010. Owing to the high flash rates, the single SEDs sometimes cannot be counted. Instead, we give the number of SED pixels, which is the same as the number of single time–frequency measurements. Previous lightning storms from 35° S (ref. 5) showed SED numbers more than an order of magnitude smaller for similar distances. **b**, The distance of Cassini to Saturn's centre in units of Saturn radii (R_S); these data help interpretation of **a**, as the signal intensities of SEDs decrease with distance squared. The peak in SED numbers on 20/21 December coincides with Cassini's closest approach to Saturn.

axisymmetric. SED-related storms were much more common in the years around Saturn's equinox. From November 2007 until now (May 2011) they were present for $\sim 69\%$ of the time, which is about 10 times more compared to the first years of Cassini's orbit around Saturn.

On 5 December the storm was located at latitude $\sim 32^{\circ}$ N, where the wind speed with respect to the Voyager SLS is nearly zero. In

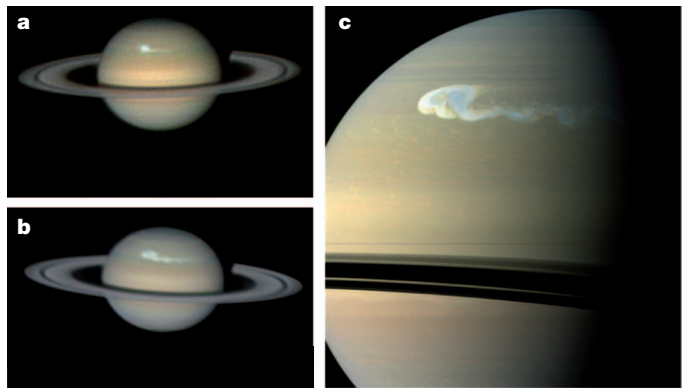


Figure 3 | Images of Saturn with the storm. **a**, Image taken on 13 December 2010 with an 11-inch Schmidt-Cassegrain telescope in the Philippines (by C.G.). The centre of the storm was located at longitude $\sim 262^{\circ}$ W and latitude $\sim 34^{\circ}$ N. At that time, the latitudinal size of the storm had already grown from 1,300 km to $\sim 6,800$ km ($\sim 7^{\circ}$). **b**, Image taken on 22 December 2010 with a 16-inch Newtonian telescope in Australia (by A.W.). It shows that the plume had grown to a latitudinal and longitudinal extent of $\sim 9,000$ km and $\sim 15,000$ km, respectively. The centre was at longitude $\sim 283^{\circ}$ W. **c**, Image taken on 24 December 2010 with the wide-angle camera of Cassini. Here the storm's latitudinal and longitudinal extent is $\sim 10,000$ km and $\sim 17,000$ km, respectively, with an eastward tail that extends much further. The storm centre was at longitude $\sim 288^{\circ}$ W. Image credits: C.G., A.W., NASA/JPL/SSI.

the interval from 5 to 8 December, the SEDs were observed at nearly constant sub-spacecraft west longitudes, indicating no drift. However, the images after 8 December showed that the core region of the storm had moved to a latitude of 34° N, where the wind speed has a local minimum around -22 m s $^{-1}$ (corresponding to a westward drift of $\sim 2.3^{\circ}$ per Earth day) and is therefore the centre of a westward jet^{7,15}. From the Earth-based images of December and early January, we derived a westward drift of $2.4^{\circ} \pm 0.1^{\circ}$ per day for the approximate centre of the storm. In the Cassini images, the western edge showed a higher drift of $2.8^{\circ} \pm 0.1^{\circ}$ per day. All images in Fig. 3 show a storm with a head on its western side and a tail to the east. As the head moves

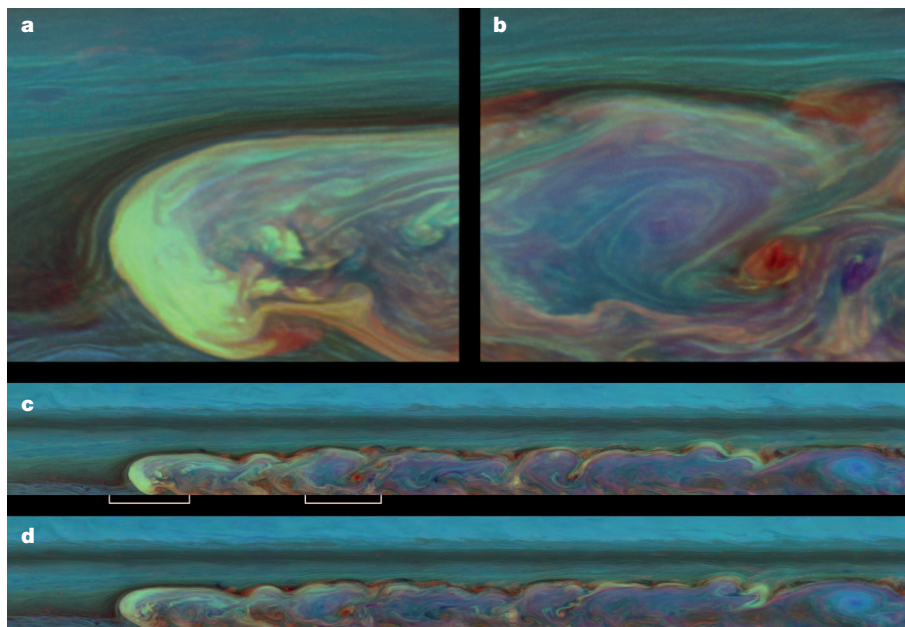


Figure 4 | False-colour views showing the height of the storm clouds. Images with three filters sensitive to different amounts of absorption by methane gas were superposed to make these colour mosaics. The filtered image at 889 nm is projected as blue, and sees only the highest clouds. The filtered image at 727 nm is projected as green, and sees high and intermediate clouds

but not the lowest clouds. The filtered image at 750 nm is projected as red, and sees clouds at all levels. **a**, **b**, Magnified views at full resolution of the areas indicated by white brackets in **c**. **d**, Image showing the storm ~ 11 h after image **c** was taken. The images were taken by Cassini's narrow angle camera on 26 February 2011 from a distance of 2.4×10^6 km. Image credit: NASA/JPL/SSI.

with the westward jet, the relative flow at lower latitudes is eastward, which accounts for the eastward tail to the south of the storm. The head plus tail had a longitudinal extent of $\sim 90^\circ$ in the late December images, and the disturbance was wrapped around the planet by February. The RPWS instrument occasionally observed SEDs over an extended longitude range. This suggests at least one small and temporarily active SED storm cell in the tail, whereas the main lightning activity took place in the head of the storm. Figure 4 shows Cassini high-resolution views of the storm in filters that bring out differences in cloud heights. Yellows and white identify optically thick clouds at high altitudes located mostly around the head of the storm.

Saturn's lightning storms are thought to be located in the water cloud layer at pressure levels around 8–10 bar (ref. 2). Optical flash observations around Saturn's vernal equinox in August 2009 have restricted the flash depth to 125–250 km below the upper cloud layer¹⁶. As most of the sunlight is absorbed around the 1-bar level¹⁷, the lightning storms are probably driven by internal heat. Moist convective plumes¹⁸ transport particles from the ammonia and ammonium hydrosulphide clouds, as well as particles produced by the lightning itself, to levels above 1 bar (ref. 19). We estimate the contribution of the current storm to Saturn's emitted power using the accumulated out-flow volume of the storm's anvil cloud. Assuming the molar mixing ratio of the water vapour in the plume before it rises is five times the solar value²⁰ ($O/H = 0.513 \times 10^{-3}$), condensation of all water would produce a temperature increase of 7.7 K. This temperature change would raise the power radiated at the latitude of the storm by 36%. Variations of this magnitude have occurred in the past. For instance, from 2004 to 2009, the emitted power was on average 17% higher in the southern hemisphere than in the northern hemisphere²¹. We assume the heated layer is 1 bar thick ($\sim 10^4 \text{ kg m}^{-2}$), because that is the depth of the clouds we are seeing. Then it would take 6–10 years to radiate away the 7.7 K temperature rise. The excess energy in such a layer, if it were 10,000 km wide and $\sim 360^\circ$ in longitude, would be comparable to the total internal heat radiated by the planet in a year, assuming the internal power is (0.78/1.78) times the total radiated power²². Thus giant storms like this might be a significant term in the internal heat budget of the planet.

The switch in hemisphere of the lightning storms could be related to the change of seasons^{7,23}. On Saturn, the seasonal effect is exaggerated by the shadow of the rings. The last SED storm in the southern hemisphere ended in mid-July 2010, ~ 11 months after equinox, and this first storm in the northern hemisphere started ~ 16 months after equinox. Convective storms around 35° N were imaged shortly after northern vernal equinox by the cameras of Voyager 1 and 2^{24,25}. SEDs were also present at those times, but they were thought to be related to an equatorial thunderstorm^{26,27}. There are similarities and differences between lightning storms on Saturn and on Earth. Saturnian thunderstorms last longer, are much larger, and occur only at very specific latitudes. SEDs are more intense than terrestrial flashes, but their duration is similar². Charging mechanisms in the water clouds of Saturnian and terrestrial thunderstorms are probably similar, and lightning on Earth also has a seasonal dependence²⁸.

Received 10 February; accepted 17 May 2011.

- Warwick, J. W. *et al.* Planetary radio astronomy observations from Voyager 1 near Saturn. *Science* **212**, 239–243 (1981).
- Fischer, G. *et al.* Atmospheric electricity at Saturn. *Space Sci. Rev.* **137**, 271–285 (2008).
- Porco, C. C. *et al.* Cassini imaging science: initial results on Saturn's atmosphere. *Science* **307**, 1243–1247 (2005).
- Dyudina, U. A. *et al.* Lightning storms on Saturn observed by Cassini ISS and RPWS during 2004–2006. *Icarus* **190**, 545–555 (2007).
- Fischer, G. *et al.* Analysis of a giant lightning storm on Saturn. *Icarus* **190**, 528–544 (2007).
- Sánchez-Lavega, A. Motions in Saturn's atmosphere: observations before Voyager encounters. *Icarus* **49**, 1–16 (1982).
- Sánchez-Lavega, A. *et al.* Deep winds beneath Saturn's upper clouds from a seasonal long-lived planetary-scale storm. *Nature* doi:10.1038/nature10203 (this issue).
- Gurnett, D. A. *et al.* The Cassini Radio and Plasma Wave Science investigation. *Space Sci. Rev.* **114**, 395–463 (2004).
- Porco, C. C. *et al.* Cassini Imaging Science: instrument characteristics and anticipated scientific investigations at Saturn. *Space Sci. Rev.* **115**, 363–497 (2004).
- Kaiser, M. L., Desch, M. D. & Connerney, J. E. P. Saturn's ionosphere: inferred electron densities. *J. Geophys. Res.* **89**, 2371–2376 (1984).
- Fischer, G., Gurnett, D. A., Zarka, P., Moore, L. & Dyudina, U. A. Peak electron densities in Saturn's ionosphere derived from the low-frequency cutoff of Saturn lightning. *J. Geophys. Res.* **116**, A04315, doi:10.1029/2010JA016187 (2011).
- Moore, L. *et al.* Cassini radio occultations of Saturn's ionosphere: model comparisons using a constant water flux. *Geophys. Res. Lett.* **33**, L22202, doi:10.1029/2006GL027375 (2006).
- Connerney, J. E. P. & Waite, J. H. New model of Saturn's ionosphere with an influx of water from the rings. *Nature* **312**, 136–138 (1984).
- Fischer, G., Gurnett, D. A., Lecacheux, A., Macher, W. & Kurth, W. S. Polarization measurements of Saturn electrostatic discharges with Cassini/RPWS below a frequency of 2 MHz. *J. Geophys. Res.* **112**, A12308, doi:10.1029/2007JA012592 (2007).
- Ingersoll, A. P., Beebe, R. F., Conrath, J. & Hunt, G. E. in *Saturn* (eds Gehrels, T. & Matthews, M. S.) 195–238 (Univ. Arizona Press, 1984).
- Dyudina, U. A. *et al.* Detection of visible lightning on Saturn. *Geophys. Res. Lett.* **37**, L09205, doi:10.1029/2010GL043188 (2010).
- Pérez-Hoyos, S. & Sánchez-Lavega, A. Solar flux in Saturn's atmosphere: maximum penetration and heating rates in the aerosol and cloud layers. *Icarus* **180**, 368–378 (2006).
- Hueso, R. & Sánchez-Lavega, A. A three-dimensional model of moist convection for the giant planets II: Saturn's water and ammonia moist convective storms. *Icarus* **172**, 255–271 (2004).
- Baines, K. H. *et al.* Storm clouds on Saturn: lightning-induced chemistry and associated materials consistent with Cassini/VIMS spectra. *Planet. Space Sci.* **57**, 1650–1658 (2009).
- Atreya, S. K. in *Galileo's Medicean Moons — Their Impact on 400 Years of Discovery* (eds Barbieri, C. *et al.*) Ch. 16 (Cambridge Univ. Press, 2010).
- Li, L. *et al.* Saturn's emitted power. *J. Geophys. Res.* **115**, E11002, doi:10.1029/2010JE003631 (2010).
- Hanel, R. A., Conrath, B. J., Kunde, V. G., Pearl, J. C. & Pirraglia, J. A. Albedo, internal heat flux, and energy balance of Saturn. *Icarus* **53**, 262–285 (1983).
- Sánchez-Lavega, A. *et al.* The Great White Spot and disturbances in Saturn's equatorial atmosphere during 1990. *Nature* **353**, 397–401 (1991).
- Hunt, G. E., Godfrey, D., Müller, J.-P. & Barrey, R. F. T. Dynamical features in the northern hemisphere of Saturn from Voyager 1 images. *Nature* **297**, 132–134 (1982).
- Sromovsky, L. A., Revercomb, H. E., Krauss, R. J. & Suomi, V. E. Voyager 2 observations of Saturn's northern mid-latitude cloud features: morphology, motions, and evolution. *J. Geophys. Res.* **88**, 8650–8666 (1983).
- Kaiser, M. L., Connerney, J. E. P. & Desch, M. D. Atmospheric storm explanation of saturnian electrostatic discharges. *Nature* **303**, 50–53 (1983).
- Zarka, P. & Pedersen, B. M. Statistical study of Saturn electrostatic discharges. *J. Geophys. Res.* **88**, 9007–9018 (1983).
- Christian, H. *et al.* Global frequency and distribution of lightning as observed from space by the Optical Transient Detector. *J. Geophys. Res.* **108** (D1), 4005, doi:10.1029/2002JD002347 (2003).

Acknowledgements G.F. was supported by the Austrian Science Fund (FWF). Cassini research at the University of Iowa was funded by NASA/JPL. We thank A. Sánchez-Lavega for launching an alert on the webpage of the Planetary Virtual Observatory and Laboratory (<http://www.pvol.uhu.es>), and the amateur astronomers who subsequently observed the storm on Saturn, namely, T. Akutsu, T. Barry, J. Castella, D. Chang, D. Gray, J. B. Jovani, W. Kivits, T. Kumamori, F. J. Melillo, D. Parker, D. Peach, J. Phillips, J.-J. Poupeau, J. Sussenbach, K. Yunoki and S. Walker. S. Ghomizadeh and T. Ikemura were the first to observe the storm before the alert.

Author Contributions G.F. analysed Cassini RPWS data and wrote the paper. W.S.K., D.A.G. and P.Z. helped in this analysis. U.A.D., A.P.I., S.P.E. and C.C.P. analysed the Cassini ISS image and calculated the energy of the storm. A.W. and C.G. imaged Saturn from the ground, and M.D. measured the size and drift of the storm from several images. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to G.F. (georg.fischer@oeaw.ac.at).

Attosecond control of electrons emitted from a nanoscale metal tip

Michael Krüger^{1*}, Markus Schenk^{1*} & Peter Hommelhoff¹

Attosecond science is based on steering electrons with the electric field of well controlled femtosecond laser pulses¹. It has led to the generation of extreme-ultraviolet pulses² with a duration of less than 100 attoseconds (ref. 3; 1 as = 10^{-18} s), to the measurement of intramolecular dynamics (by diffraction of an electron taken from the molecule under scrutiny^{4,5}) and to ultrafast electron holography⁶. All these effects have been observed with atoms or molecules in the gas phase. Electrons liberated from solids by few-cycle laser pulses are also predicted^{7,8} to show a strong light-phase sensitivity, but only very small effects have been observed¹⁴. Here we report that the spectra of electrons undergoing photoemission from a nanometre-scale tungsten tip show a dependence on the carrier-envelope phase of the laser, with a current modulation of up to 100 per cent. Depending on the carrier-envelope phase, electrons are emitted either from a single sub-500-attosecond interval of the 6-femtosecond laser pulse, or from two such intervals; the latter case leads to spectral interference. We also show that coherent elastic re-scattering of liberated electrons takes place at the metal surface. Owing to field enhancement at the tip, a simple laser oscillator reaches the peak electric field strengths required for attosecond experiments at 100-megahertz repetition rates, rendering complex amplified laser systems dispensable. Practically, this work represents a simple, extremely sensitive carrier-envelope phase sensor, which could be shrunk in volume to about one cubic centimetre. Our results indicate that the attosecond techniques developed with (and for) atoms and molecules can also be used with solids. In particular, we foresee subfemtosecond, subnanometre probing of collective electron dynamics (such as plasmon polaritons⁹) in solid-state systems ranging in scale from mesoscopic solids to clusters and to single protruding atoms.

A nanoscale solid-state system is also of interest for applications. Steering electrons with the force exerted by a synthesized few-cycle light field is predicted to allow ultimate speeds to be reached in electronics: that is, up to optical frequencies, with a typical timescale of femtoseconds ('lightwave electronics'¹⁰, so called by analogy to microwave electronics with semiconductor chips). However, typical electron energies in conventional electronics lie in the range of a few electron volts, corresponding to a free-electron velocity of $\sim 1 \text{ nm fs}^{-1}$. Thus, the possible speed-up to optical frequencies in electronics can only be achieved if a nanometre-scale solid-state system is used⁸. Because the electron current is switched on and off by the light field, such a device could be called an optical attosecond field-effect transistor. The present work is, to our knowledge, the first research along these lines.

Our experiment centres on a metal tip of radius $\sim 10\text{--}20 \text{ nm}$, the apex of which lies in the focus of a linearly polarized laser beam that consists of few-cycle laser pulses (Fig. 1). The tip is beneficial for two reasons. First, electric field enhancement takes place because of its sharpness¹¹. At the tip's apex, the electric field is around five times higher than in the laser focus alone; this corresponds to a ~ 25 -fold increase in intensity. A laser oscillator is therefore sufficient to reach the regime where Newton's equations of motion describe the electron's

response quite well; this marks the regime of attosecond science¹. Field enhancement near nanostructures has also facilitated the generation of high-harmonic radiation from gas atoms with an oscillator only¹², whereas in conventional gas phase experiments amplified laser systems are used¹³. Second, because of the localized nature of field enhancement, electron emission is limited to a single well-defined site with a diameter of $\sim 10 \text{ nm}$, right on the tip's apex. As this is much smaller than the focal diameter of the laser beam, the laser intensity can be well approximated as constant over the electron emission area. Thus electrons are emitted from a single nanometre-sized area exposed to a well-defined laser intensity.

This condition does not usually hold for photoemission from plane surfaces, where local laser intensities differ greatly, owing to the Gaussian profile of the laser beam and hotspot effects due to potential surface roughness. We conjecture that carrier-envelope phase effects are therefore blurred when the electron emission current from a larger area is measured. Presumably for these reasons, only minuscule carrier-envelope phase effects have so far been observed in nonlinear photoemission from a gold cathode, even though extremely short near-infrared laser pulses (with a duration of 4 fs) were used¹⁴. A previous attempt to measure a carrier-envelope phase dependence in photoemission from sharp tips lacked spectral information, which presumably hindered detection of the signal¹⁵. An attosecond streaking experiment from a solid has been reported¹⁶, but there the electrons were photoemitted from the metal surface by extreme-ultraviolet pulses before they interacted with the infrared light field, thereby mitigating hotspot effects. Collective electron effects¹⁷ and strong-field effects^{18,19} have been observed in tip-enhanced electron emission, while ref. 20 shows that thermal effects can be ruled out here.

In the present experiment, we focus ~ 6 -fs pulses tightly on a tungsten tip; the pulses are derived from a Ti:sapphire oscillator with a stabilized

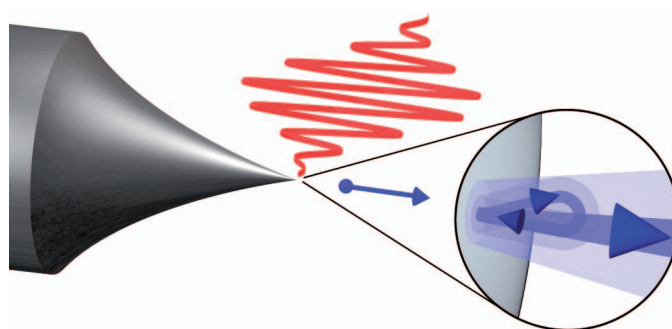


Figure 1 | Overview of the experiment. Electrons, indicated in blue, are emitted (large blue arrow pointing away from surface) from a sharp tungsten tip (left) by, and interact with, a few-cycle laser electric field (red waveform). Controlled by the carrier-envelope phase, the liberated electron may be driven back into the tip, where it can scatter elastically and gain more energy in the laser field before it reaches the detector (not shown).

¹Ultrafast Quantum Optics Group, Max-Planck-Institut für Quantenoptik, Hans-Kopfermann-Straße 1, D-85748 Garching bei München, Germany.

*These authors contributed equally to this work.

carrier-envelope phase and an 80-MHz repetition rate (Fig. 1; the set-up is described in more detail in ref. 19 and in Supplementary Information). A small negative extraction voltage is applied to the tip, resulting in a d.c. electric field strength of $\sim 0.4 \text{ GV m}^{-1}$ at the tip's apex. We record photoelectron spectra with a retarding field spectrometer.

Conclusions can be drawn about the dynamics of the photoelectrons from the overall shape of the spectra. In Fig. 2a, we present a carrier-envelope phase averaged spectrum obtained with 240-pJ pulse energy (a peak intensity of $\sim 4 \times 10^{11} \text{ W cm}^{-2}$ in the bare focus), with an average yield of about one electron per pulse. The spectrum is governed by above-threshold photoemission peaks spaced approximately by the photon energy (1.56 eV) on top of an overall exponential decay¹⁹. This decay is followed by a plateau, a region of almost constant count rate extending from $\sim 4.5 \text{ eV}$ to a soft cut-off located at $\sim 13 \text{ eV}$. The appearance of the plateau indicates that coherent elastic re-scattering of electrons is taking place. This effect is well known in the ionization of atoms in the gas phase: a small fraction of the photoelectrons is driven back into the tip by the laser field, scatters elastically off the tip, and gains more energy in the laser field before being detected²¹. Models presented below strongly support this notion. A more detailed investigation of electron re-scattering dynamics from a metal is currently under way. We note that recombination of the active electron can lead to emission of high-harmonic radiation¹. We expect this process to also take place at tips.

The electric field of the laser pulses can be written as $E(t) = f(t)\cos(\omega t + \phi_{\text{CE}})$, with $f(t)$ describing the pulse envelope, ω the laser's centre (circular) frequency, and ϕ_{CE} the carrier-envelope phase. Figure 2b shows a contour plot of individual electron spectra as a function of the carrier-envelope phase offset, which is given by the sum of ϕ_{CE} and a constant experimental phase difference to be

determined by theory (set to 0 here, see below). Clearly, the spectral features are strongly modulated with the carrier-envelope phase: both maxima and minima show pronounced modulation effects with a period of 2π . In Fig. 2a (green points) we display the modulation depth of the count rate for different energy positions (for a definition, see Supplementary Information). At low energy, the modulation depth amounts to several per cent and gains strength in the plateau (10–25%). In the region of the 13-eV cut-off, it increases to $\sim 100\%$ and here the carrier-envelope phase almost completely determines if a photoelectron will be detected.

The visibility of the plateau peaks is particularly strongly affected by the carrier-envelope phase. Figure 2c depicts individual spectra for four carrier-envelope phase settings spaced by $\pi/2$. It is evident that for certain phases peaks are clearly visible, whereas for others the peaks almost completely disappear. An analysis of the average peak visibility is shown in Fig. 2d (for details, see Supplementary Information). It is approximately sinusoidally modulated with the carrier-envelope phase, ranging from $\sim 10\%$ for $\phi_{\text{CE}} \approx -0.2\pi$ to $\sim 30\%$ for $\phi_{\text{CE}} \approx 0.8\pi$. We will show below that the peaks arise from quantum mechanical interference of electron wave packets re-scattering at the tip in different optical cycles. The visibility can be identified as the degree of spectral interference. Strong interference indicates that (at least) two wave-packet components contribute to the plateau. In contrast, the absence of interference implies that only a single electron wave packet from one optical cycle contributes.

In addition, we observe that the position of the high-energy cut-off changes with carrier-envelope phase, as shown in Fig. 2b (red line). It varies between 12.3 eV and 13.6 eV. Notably, the behaviour of peak visibility and cut-off position is maximally out of phase: the phase difference amounts to $\pi + (80 \pm 160) \text{ mrad}$.

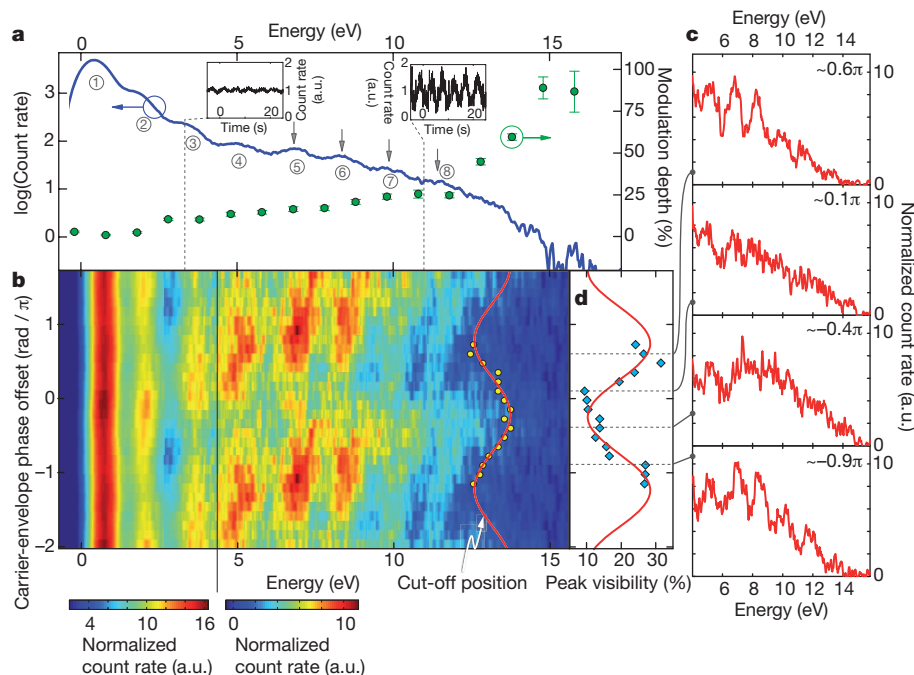


Figure 2 | Carrier-envelope phase modulation in photoelectron spectra.

a, Carrier-envelope phase-averaged electron count rate as function of energy (blue solid curve). About three photon orders, indicated by encircled numbers, are visible in the direct part (that is, when the electrons do not re-collide with the tip; $E = 0$ to $\sim 4.3 \text{ eV}$). For $E > 4.5 \text{ eV}$, the plateau region starts with five more photon orders visible. The green points depict the modulation depth of the count rate when varying the carrier-envelope phase (error bars, fit errors of modulation curves; see Supplementary Information). Insets, carrier-envelope phase modulation in the photocurrent with the spectrometer acting as an energy high-pass filter at 3 eV (left inset) and 11 eV (right inset; both with the carrier-envelope offset frequency set to $f_{\text{CEO}} = \phi_{\text{CE}}/2\pi = 0.2 \text{ Hz}$). a.u., arbitrary units. **b**, Contour plot of the electron count rate as function of carrier-envelope phase

offset and energy. ϕ_{CE} was increased in steps of $\pi/8$. At 4.3 eV, the plot is split into two regions for better visibility (see colour scales under). In each region, we plot the normalized count rate (Supplementary Information). Measured data range over 2π and are extended over $\sim 4\pi$ for better visibility. Yellow circles show the position of the cut-off for a given carrier-envelope phase offset (red curve, sinusoidal fit). **c**, Individual electron spectra extracted from the contour plot in **b** for four different carrier-envelope phase offsets separated by $\pi/2$. Only the plateau region is shown. Fringes are clearly visible for 0.6π and -0.9π , but almost no fringes are visible for 0.1π and -0.4π . **d**, Blue diamonds, average peak visibility in the plateau region (red curve, sinusoidal fit). The peaks used to determine the visibility are marked with grey arrows in **a**. Note that peak visibility and cut-off position are nearly maximally out of phase.

We interpret our experimental findings with the aid of two theoretical models. The first model used is the semiclassical Simple Man's Model (SMM)²². In brief, an electron is liberated by optically induced tunnelling and subsequently propagates in the laser electric field on classical trajectories. The model has been extended to account for the matter-wave nature of liberated electrons by including the accumulated quantum mechanical phase of the corresponding wave packets²³. Trajectories with different start times within the pulse that lead to the same final energy interfere, resulting in interference structures in the energy domain²⁴.

In Fig. 3a we present spectra simulated by the SMM that are similar to those in Fig. 2b; we assume a 6.3-fs pulse with a peak electric field of 10.4 GV m^{-1} . All the main features of the experimental data are qualitatively reproduced, notably the shift of the cut-off position and the correlated change in peak visibility. A region is observed where no spectral interference occurs, centred around $\phi_{\text{CE}} \approx 0$, corresponding to a 'cosine-like' pulse. Simultaneously, the cut-off position is located at the highest energy. Figure 3b illustrates the physical origin of both effects (namely, no spectral interference and the cut-off position located at the highest energy) for $\phi_{\text{CE}} \approx 0$: only a single trajectory from one optical half-cycle, reaching the highest possible kinetic energy,

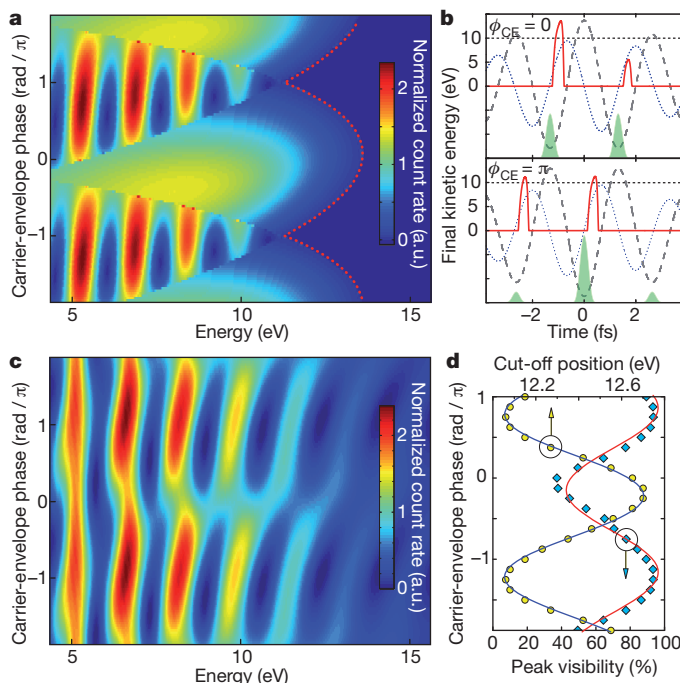


Figure 3 | Theoretical modelling of the experimental data. **a**, Contour plot representation of the count rate as a function of energy and carrier-envelope phase, according to the extended SMM. Here only re-scattered electrons are considered (only those contribute to plateau energies). Regions with quantum mechanical interference fringes ($\phi_{\text{CE}} \approx \pm \pi$) and without ($\phi_{\text{CE}} \approx 0$) are evident (sharp borders reflect the semiclassical nature of the model). Furthermore, the highest electron energies are reached for carrier-envelope phase settings corresponding approximately to the most featureless electron spectrum ($\phi_{\text{CE}} \approx 0$). Dotted red curve, the classical cut-off position. **b**, Final kinetic energy (red curve) and emission probability (green area) as a function of emission time for phases 0 and π . The emission probability in the model follows the electric field (dashed grey curve). Emission occurs only when the electric field is negative and points into the metal, thus pulling electrons out of the metal. The electron's kinematics can be inferred more easily from the light field's vector potential $A(t)$, shown with a blue dotted line ($E(t) = -\partial A(t)/\partial t$). Notably, either one or two trajectories contribute significantly to the kinetic energy of, for example, 10 eV (marked by the dotted horizontal lines). Within each cycle a second trajectory also exists, but the corresponding emission probability can be neglected with respect to the dominating one. **c**, As in **a** but calculated by numerically integrating the TDSE. **d**, Average peak visibility and cut-off position from **c**.

yields a significant contribution to the electron count rate in the high-energy part of the spectrum. In contrast, the region around $\phi_{\text{CE}} \approx \pm \pi$, which corresponds to a 'minus-cosine-like' pulse, exhibits the strongest peak structure and the lowest cut-off energy. Two trajectories from subsequent cycles contribute here, with the times of both emission and re-scattering differing approximately by the duration of an optical cycle, T_{opt} . This translates into interference fringes in the energy domain, with a spacing of $\Delta E \approx h/T_{\text{opt}} = h/(2.67 \text{ fs}) \approx 1.56 \text{ eV}$, approximately equalling the photon energy²³. A similar behaviour was demonstrated for strong-field photoemission from atomic gases²⁴, and also in ref. 25, where the generation of single subfemtosecond extreme-ultraviolet radiation bursts was deduced from the absence of interference in the cut-off region of the extreme-ultraviolet spectrum. Accordingly, single electron matter-wave packets with initially subfemtosecond emission duration (Fourier limit of cut-off part, $\sim 450 \text{ as}$) can be generated from a metallic tip by appropriate setting of the carrier-envelope phase. Moreover, the timing of the emission and propagation of electron wave packets undergoing re-scattering near the metal surface can be controlled with a precision of $\sim 80 \text{ as}$ by changing the carrier-envelope phase, as inferred from the error of the phase offset in the sinusoidal fits to cut-off and visibility. The presence of interference indicates the coherent nature of the re-scattering process off a metal surface, which has not been reported before.

The second model, a fully quantum mechanical treatment, shows good quantitative agreement with the experiment. In Fig. 3c we present energy spectra retrieved from a numerical integration of the one-dimensional time-dependent Schrödinger equation (TDSE)¹⁵. The parameters were adapted to match the experimental data (see Supplementary Information for details). The average peak visibility and the cut-off position are depicted in Fig. 3d and behave similarly to the experimental data. We have shifted the carrier-envelope phase axis of the experimental data so that the maxima of the cut-off position curves coincide, thereby zeroing the experimental phase offset. Thus, the pulse is cosine-like ($\phi_{\text{CE}} = 0$) for a carrier-envelope phase offset of $(0.00 \pm 0.05)\pi$. The spectral shift of the cut-off with carrier-envelope phase in the TDSE model is a weaker effect (peak-to-peak shift $\sim 0.7 \text{ eV}$) than in the experimental data (peak-to-peak shift $\sim 1.3 \text{ eV}$, compare Fig. 3d and Fig. 2b), mainly because strong smoothing had to be applied to the TDSE spectra in order to suppress effects caused by the pronounced peak structure. The SMM (Fig. 3a), in contrast, reveals a larger shift (peak-to-peak $\sim 2.3 \text{ eV}$), mainly due to its semiclassical nature. Given the simplicity of the models, the agreement is satisfactory.

The peak visibility is lower in the experiment than in both theoretical treatments, which we attribute to the spectrometer resolution and low counting statistics. Moreover, in both models we consider only a single initial electronic state at the Fermi level. A metal, however, comprises many populated states with a large spread of initial energies. In future work, a more elaborate simulation^{7,8} will be needed to fully take this into account, together with possible plasmonic effects; the technique could then be used to draw conclusions about the underlying subfemtosecond dynamics.

Three points are noteworthy. First, the observable carrier-envelope phase of the enhanced field at the tip's surface should be phase-shifted with respect to the phase of the driving laser field, owing to the plasmonic response of the metal^{8,12}. Spectra measured from a tip with strong plasmonic behaviour (such as one made from silver) would thus allow information to be obtained about the collective electron response on the subnanometre–subfemtosecond scale. Second, in the SMM, an emission process according to optically induced tunnelling has been assumed, with the electrons' classical trajectories starting at the tunnel exit with zero initial momentum^{6,26} and with an emission probability modelled along the lines of the ADK rate (see Supplementary Information). We also tried other emission processes (non-adiabatic tunnelling²⁷, multiphoton photoemission) but found the best agreement using tunnelling, although here the Keldysh parameter is

~2. We stress that in this parameter range, the emission process encompasses and cannot be separated from strong-field effects after the electron has been classically emitted. Third, although the exact quantum dynamics of electrons in this new system is complex and to understand it fully will require much further investigation, it has become clear from experiments with neutral atoms²¹ and negatively charged ions²⁸ that the re-scattering scheme and the concomitant tell-tale plateau seem to be universal, in the sense that they exist despite qualitatively different potentials (long²¹ versus short range²⁸). Also, we note that related initial theoretical work on photoemission from metals contains hints in this direction²⁹. Our experimental results provide strong evidence for very similar underlying physics, even though here the dimensions involved are very different from those involved in studies on atoms or ions: the electron source and scatterer, namely the solid tip, is much larger than the classical oscillation amplitude of the electron in the laser field (a few ångströms). It will be interesting to investigate the implications of this. For example, does scattering take place at the extended surface or at individual surface atoms—and does this depend on the material and its orientation? Angle-resolved spectra might yield information.

With longer laser wavelengths, the energy of the re-colliding electron increases and can surpass several tens of electron volts. Hence, new forms of (time-resolved) surface science techniques—such as low-energy electron diffraction—with electrons originating from and probing the surface, might come into reach, with typical timescales of 100 as. Extending this work towards more complex, lithographically grown nanoscale objects^{11,30} will pave the way to lightwave electronics, where the electric field of light steers electrons; for example, the area between a source and a drain electrode could be switched between conducting and insulating states. Also, a simple stand-alone die-sized (1-cm³) sensor device for the carrier-envelope phase might be feasible: such a device could be extremely sensitive, and comprise only a tip, a retardation grid and an electron multiplier.

Received 22 February; accepted 12 May 2011.

- Corkum, P. B. & Krausz, F. Attosecond science. *Nature Phys.* **3**, 381–387 (2007).
- Antoine, P., L'Huillier, A. & Lewenstein, M. Attosecond pulse trains using high-order harmonics. *Phys. Rev. Lett.* **77**, 1234–1237 (1996).
- Goulielmakis, E. *et al.* Single-cycle nonlinear optics. *Science* **320**, 1614–1617 (2008).
- Niikura, H. *et al.* Sub-laser-cycle electron pulses for probing molecular dynamics. *Nature* **417**, 917–922 (2002).
- Baker, S. *et al.* Probing proton dynamics in molecules on an attosecond time scale. *Science* **312**, 424–427 (2006).
- Huismans, Y. *et al.* Time-resolved holography with photoelectrons. *Science* **331**, 61–64 (2011).
- Lemell, C., Tong, X.-M., Krausz, F. & Burgdörfer, J. Electron emission from metal surfaces by ultrashort pulses: determination of the carrier-envelope phase. *Phys. Rev. Lett.* **90**, 076403 (2003).
- Stockman, M. I. & Hewageegana, P. Absolute phase effect in ultrafast optical responses of metal nanostructures. *Appl. Phys. A* **89**, 247–250 (2007).
- Stockman, M. I. Nanofocusing of optical energy in tapered plasmonic waveguides. *Phys. Rev. Lett.* **93**, 137404 (2004).
- Goulielmakis, E. *et al.* Attosecond control and measurement: lightwave electronics. *Science* **317**, 769–775 (2007).
- Novotny, L. & van Hulst, N. Antennas for light. *Nature Photon.* **5**, 83–90 (2011).
- Kim, S. *et al.* High-harmonic generation by resonant plasmon field enhancement. *Nature* **453**, 757–760 (2008).
- Bucksbaum, P. H. The future of attosecond spectroscopy. *Science* **317**, 766–769 (2007).
- Apolonski, A. *et al.* Observation of light-phase-sensitive photoemission from a metal. *Phys. Rev. Lett.* **92**, 073902 (2004).
- Cavalieri, A. L. *et al.* Ultrafast electron pulses from a tungsten tip triggered by low-power femtosecond laser pulses. *Phys. Rev. Lett.* **97**, 247402 (2006).
- Cavalieri, A. L. *et al.* Attosecond spectroscopy in condensed matter. *Nature* **449**, 1029–1032 (2007).
- Yanagisawa, H. *et al.* Optical control of field-emission sites by femtosecond laser pulses. *Phys. Rev. Lett.* **103**, 257603 (2009).
- Bormann, R., Gulde, M., Weismann, A., Yalunin, S. V. & Ropers, C. Tip-enhanced strong-field photoemission. *Phys. Rev. Lett.* **105**, 147601 (2010).
- Schenk, M., Krüger, M. & Hommelhoff, P. Strong-field above-threshold photoemission from sharp metal tips. *Phys. Rev. Lett.* **105**, 257601 (2010).
- Kealhofer, C., Foreman, S. M., Gerlich, S. & Kasevich, M. A. Ultrafast laser-triggered emission from hafnium carbide tips. *Phys. Rev. B* (submitted); preprint at (<http://arxiv.org/abs/1104.1452>) (2011).
- Paulus, G. G., Nicklich, W., Xu, H. L., Lambropoulos, P. & Walther, H. Plateau in above-threshold ionization spectra. *Phys. Rev. Lett.* **72**, 2851–2854 (1994).
- Corkum, P. B. Plasma perspective on strong-field multiphoton ionization. *Phys. Rev. Lett.* **71**, 1994–1997 (1993).
- Milošević, D. B., Paulus, G. G., Bauer, D. & Becker, W. Above-threshold ionization by few-cycle pulses. *J. Phys. B* **39**, R203–R262 (2006).
- Lindner, F. *et al.* Attosecond double-slit experiment. *Phys. Rev. Lett.* **95**, 040401 (2005).
- Baltuška, A. *et al.* Attosecond control of electronic processes by intense light fields. *Nature* **421**, 611–615 (2003).
- Zherebtsov, S. *et al.* Controlled near-field enhanced electron acceleration from dielectric nanospheres with intense few-cycle laser fields. *Nature Phys.* advance online publication, doi:10.1038/nphys1983 (24 April 2011).
- Yudin, G. L. & Ivanov, M. Y. Nonadiabatic tunnel ionization: looking inside a laser cycle. *Phys. Rev. A* **64**, 013409 (2001).
- Gazibegović-Busuladžić, A. *et al.* Electron rescattering in above-threshold photodetachment of negative ions. *Phys. Rev. Lett.* **104**, 103004 (2010).
- Faisal, F. H. M., Kamiński, J. Z. & Saczuk, E. Photoemission and high-order harmonic generation from solid surfaces in intense laser fields. *Phys. Rev. A* **72**, 023412 (2005).
- Aeschlimann, M. *et al.* Adaptive subwavelength control of nano-optical fields. *Nature* **446**, 301–304 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Kling, C. Lemell, G. Wachter and B. Bergues for discussions, and J. Hoffrogge for reading the text before submission. This work has been supported in part by the European Union (FP7-IRG).

Author Contributions All authors contributed to all parts of the work.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to P.H. (peter.hommelhoff@mpq.mpg.de).

Thermal spin current from a ferromagnet to silicon by Seebeck spin tunnelling

Jean-Christophe Le Breton¹, Sandeep Sharma^{1,2,3}, Hidekazu Saito², Shinji Yuasa² & Ron Jansen²

Heat generation by electric current, which is ubiquitous in electronic devices and circuits, raises energy consumption and will become increasingly problematic in future generations of high-density electronics. The control and re-use of heat are therefore important topics for existing and emerging technologies, including spintronics. Recently it was reported that heat flow within a ferromagnet can produce a flow of spin angular momentum—a spin current—and an associated voltage¹. This spin Seebeck effect has been observed in metallic^{1,2}, insulating³ and semiconductor ferromagnets⁴ with temperature gradients across them. Here we describe and report the demonstration of Seebeck spin tunnelling—a distinctly different thermal spin flow, of purely interfacial nature—generated in a tunnel contact between electrodes of different temperatures when at least one of the electrodes is a ferromagnet. The Seebeck spin current is governed by the energy derivative of the tunnel spin polarization. By exploiting this in ferromagnet–oxide–silicon tunnel junctions, we observe thermal transfer of spins from the ferromagnet to the silicon without a net tunnel charge current. The induced spin accumulation scales linearly with heating power and changes sign when the temperature differential is reversed. This thermal spin current can be used by itself, or in combination with electrical spin injection, to increase device efficiency. The results highlight the engineering of heat transport in spintronic devices and facilitate the functional use of heat.

Spin current is a central aspect of spintronics^{5,6} and is typically generated by spin-polarized charge current, through spin–orbit interaction or by magnetization dynamics. Recently, the potential of thermoelectric effects in magnetic nanostructures^{7–11} has been recognized and emphasized¹². Notably, the spin Seebeck effect has been observed in ferromagnetic materials with temperature differences between opposite ends^{1–4}. A microscopic description has been given in terms of a spin-dependent Seebeck coefficient^{1,2}, although spin current carried by magnons¹³ and mechanisms without a global spin current or spin accumulation in the ferromagnet^{4,14,15} have been proposed for some of the experiments^{1,3,4}. The Seebeck spin tunnelling (SST) we describe here is distinctly different in that it involves a thermal spin current that is created in a magnetic tunnel contact. It is associated purely with the tunnel interface. We observe it in a tunnel junction that has one ferromagnetic electrode and one non-magnetic electrode (in our case silicon) held at different temperatures, T_{FM} and T_{Si} , respectively (Fig. 1). The temperature difference, $\Delta T = T_{\text{Si}} - T_{\text{FM}}$, drives preferential tunnel transfer of electrons of either majority or minority spin from the ferromagnet to the silicon. A spin accumulation, $\Delta\mu$, is induced in the silicon, characterized by a difference in the densities and electrochemical potentials of electrons with their magnetic moments respectively parallel and antiparallel to the magnetization of the ferromagnet.

To observe Seebeck spin tunnelling, we apply a heating current with density J_{heating} to the silicon electrode (Fig. 2), causing Joule dissipation in that electrode and raising its temperature with respect to that of the ferromagnetic electrode. Measuring under the condition of zero net

tunnel charge current, I_{T} , between the ferromagnet and the silicon (Fig. 2a), this produces a voltage $V(I_{\text{T}} = 0) = V_0 + \Delta V_{\text{TH}}$. The first term includes the ohmic voltage drop over part of the heater resistance (the Cr/Au contact and the silicon) and the ordinary charge-related thermovoltage that develops across the contact to maintain $I_{\text{T}} = 0$. The second term is the SST voltage, that is, the additional voltage across the ferromagnet–oxide–silicon tunnel contact due to the thermally induced spin accumulation, $\Delta\mu$, in the silicon. It is detected by measuring the change in voltage in response to an applied magnetic field, B_z , transverse to the spins in the silicon (Hanle geometry, with the magnetic field applied along the z axis, perpendicular to both the in-plane magnetization of the ferromagnet and the injected spins). This causes spin precession and a reduction of $\Delta\mu$ to zero with a characteristic Lorentzian line shape¹⁶, also allowing it to be distinguished from any effect B_z might have on V_0 . The detected voltage change, ΔV_{TH} , is given by $\Delta V_{\text{TH}} = V(\Delta\mu) - V(\Delta\mu = 0) = \text{TSP} \times \Delta\mu/2$, where TSP is the tunnel spin polarization associated with the ferromagnet–oxide interface. The voltage V is defined as $V_{\text{Si}} - V_{\text{FM}}$, where V_{Si} and V_{FM} are the respective potentials of the silicon and ferromagnetic electrodes.

A significant thermal spin accumulation is observed in the silicon on heating it so that $T_{\text{Si}} > T_{\text{FM}}$, as evidenced by a Hanle curve (Fig. 2b) characteristic of spin polarization induced in the silicon¹⁶. The Hanle curve is identical for both directions of the Joule heating current, implying that the sign and magnitude of the spin polarization induced in the silicon are the same for both current directions. The peak amplitude scales quadratically with the heating current density (Fig. 2c) and, thus, linearly with the applied heating power (Fig. 2d, where the power, given per unit volume of silicon, equals $\rho J_{\text{heating}}^2$, with ρ the silicon's resistivity). The observed scaling is consistent with a thermally induced spin accumulation, which is expected to be linear in ΔT and J_{heating}^2 for

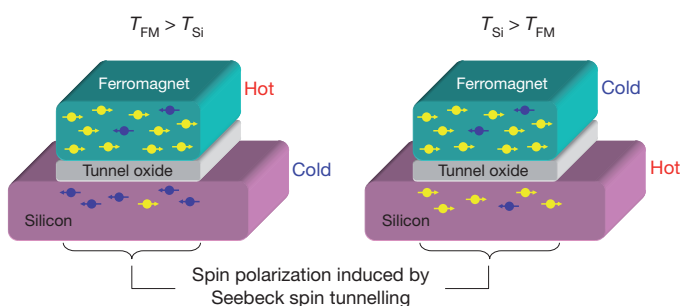


Figure 1 | Basic concept of Seebeck spin tunnelling. In a tunnel contact between a ferromagnetic electrode and a non-magnetic electrode, for instance silicon, a temperature difference between the electrodes causes a transfer of spin angular momentum from the ferromagnet to the non-magnetic electrode. The thermal spin current requires no tunnel charge current. The induced spin polarization is reversed when the temperature difference is reversed and is parallel or antiparallel to the magnetization of the ferromagnet.

¹Netherlands Foundation for Fundamental Research on Matter (FOM), 3502 GA Utrecht, The Netherlands. ²National Institute of Advanced Industrial Science and Technology (AIST), Spintronics Research Center, Tsukuba, Ibaraki 305-8568, Japan. ³Zernike Institute for Advanced Materials, Physics of Nanodevices, University of Groningen, 9747 AG, Groningen, The Netherlands.

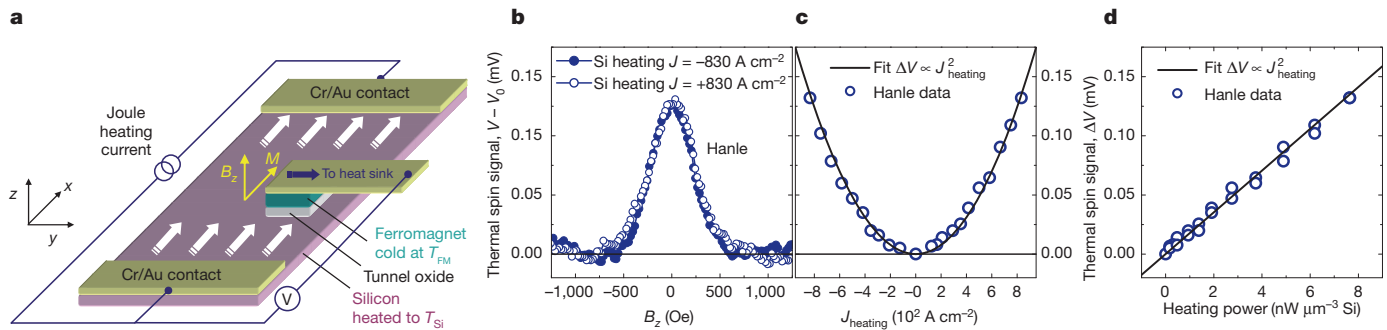


Figure 2 | Observation of thermal spin current from ferromagnet to silicon by Seebeck spin tunnelling. **a**, Device layout, showing the 3- μm -thick, 800- μm -wide, 4-mm-long silicon strip (purple), the 1.5-nm-thick $\text{SiO}_2/\text{Al}_2\text{O}_3$ tunnel barrier (light grey) and the ferromagnetic $\text{Ni}_{80}\text{Fe}_{20}$ electrode (green). The silicon is heated by Joule dissipation from a constant current applied using two Cr/Au contacts at opposite ends, resulting in $T_{\text{Si}} > T_{\text{FM}}$. The voltage across the tunnel contact is measured in the absence of a tunnel charge current. The voltage lead simultaneously acts as a thermal sink that keeps the ferromagnet cold. **b**, Thermally induced spin accumulation in silicon, detected as a voltage

change, ΔV_{TH} , in an external magnetic field, B_z , applied perpendicular to the tunnel interface (Hanle geometry). The plot shows results for heating current in two opposite directions. The magnetization (M) of the ferromagnet remains in the plane of the thin film owing to shape anisotropy. **c**, Measured ΔV_{TH} (symbols) versus heating current density, J_{heating} , together with a quadratic fit (solid line). **d**, Measured ΔV_{TH} (symbols) as a function of Joule heating power, and a linear fit (solid line). All data were measured for p-type silicon with a resistivity of 11 $\text{m}\Omega\text{ cm}$ at 300 K, for a tunnel contact area of $180\text{ }\mu\text{m} \times 400\text{ }\mu\text{m}$ and for a base temperature of 300 K.

Joule heating. Similar results are obtained with a four-terminal geometry (Supplementary Information). For the maximum heating power ($7.6\text{ nW }\mu\text{m}^{-3}$), we find that $\Delta V_{\text{TH}} = 0.13\text{ mV}$. If we assume that the TSP (30%) previously determined¹⁷ for $\text{Ni}_{80}\text{Fe}_{20}/\text{Al}_2\text{O}_3$ interfaces in metal tunnel junctions is the same in our $\text{Ni}_{80}\text{Fe}_{20}/\text{Al}_2\text{O}_3/\text{Si}$ junctions, the thermally induced $\Delta\mu$ is 0.87 meV. The SST thus produces a significant spin polarization in the silicon.

The sign of the thermal spin accumulation is determined by direct comparison with electrically induced spin accumulation measured in the same tunnel contact, driving a non-zero tunnel current across the tunnel barrier using the three-terminal Hanle method¹⁶. The electrically induced spin accumulation gives rise to a Hanle signal (Supplementary Fig. 4) with a Lorentzian line shape and width similar to that found in previous work¹⁶ and to the thermally induced signal (Fig. 2), further supporting spin accumulation as the origin of the latter. The electrically induced spin accumulation is negative for negative tunnel current and positive for positive tunnel current ($V_{\text{Si}} > V_{\text{FM}}$, extraction¹⁶ of holes from p-type silicon), and the latter, positive, spin accumulation is the same sign as the thermally induced spin accumulation (Fig. 2). Thus, thermal transfer of spin with $T_{\text{Si}} > T_{\text{FM}}$ produces a spin polarization with the same sign as that produced by electrical hole extraction. Such extraction induces in the silicon a magnetic moment parallel to the magnetization of the ferromagnet if the TSP for $\text{Ni}_{80}\text{Fe}_{20}/\text{Al}_2\text{O}_3$ interfaces is positive (dominated by majority spin electrons), as it is in metal-based tunnel junctions. Hence, the spin accumulation induced by the SST (Fig. 2; silicon heating with positive ΔV) corresponds to majority spin accumulation in the silicon (that is, a larger number of electrons with spin parallel to that of the majority spin in the ferromagnet, which is equivalent to a smaller number of holes with that spin direction).

The observed thermal spin accumulation is quadratic in the drive current, allowing us to exclude as possible origins phenomena that are not even but odd functions of current, such as the spin Hall effect. Other known magnetic or thermomagnetic effects, such as the Hall, Nernst, Ettingshausen and Righi–Leduc effects¹⁸, can be ruled out because they have a different symmetry and/or would not produce the Lorentzian magnetic field dependence characteristic of the Hanle effect (Supplementary Information). More conclusively, by measuring the Hanle signal in larger magnetic fields, for which the magnetization of the ferromagnet rotates out of the plane of the device, we performed an experimental test (Supplementary Information) that rules out all the known thermomagnetic effects. Because the field and injected spins are then no longer orthogonal, the spins no longer precess and a characteristic recovery of the spin accumulation is observed at high magnetic field values. This recovery occurs only if the spin accumulation

is produced by transfer of spins from the ferromagnet; hence, this observation excludes any source of spin accumulation that does not involve transfer from the ferromagnet. Unintentional electrical injection through shunting of part of the heating current by the tunnel contact can be ruled out for several reasons, including the signal having the wrong sign (Supplementary Information). The data in Fig. 2 thus unambiguously demonstrate thermal transfer of spin from the ferromagnet to the silicon.

For the microscopic origin of Seebeck spin tunnelling, we consider electrons as the carrier of spin across the tunnel barrier. Direct (elastic) tunnelling results in a spin current without charge current if there is a different thermal electron distribution in the two electrodes and the TSP varies with electron energy. We illustrate the case with $T_{\text{Si}} > T_{\text{FM}}$ in Fig. 3a. The excess density of electrons with energy above the Fermi level, E_{F} , in the hot silicon creates a flow of electrons tunnelling to the cold ferromagnet. Simultaneously, the excess number of empty states below the silicon Fermi level creates an electron flow in the opposite direction. Even if the total charge current is zero, there is a net spin current because the two oppositely directed tunnel currents have different TSPs. This is a consequence of the energy variation of the spin-polarized electronic structure of ferromagnetic materials around E_{F} . Indeed, for cobalt, CoFe and $\text{Ni}_{80}\text{Fe}_{20}$, it has been found that the spin polarization of electrons tunnelling through an Al_2O_3 barrier into a metal counter electrode is different for electron energies below and above E_{F} , decaying most significantly in the second case^{19,20}.

The salient features of SST are illustrated using a free-electron elastic tunnelling model, including the energy variation of the TSP as a phenomenological parameter. It also explicitly includes the spin accumulation and the feedback it has on the tunnel spin and charge currents. The thermally induced spin accumulation is thus calculated self-consistently (Supplementary Information). First, a non-zero ΔT indeed produces a spin current and, thereby, a spin accumulation in the silicon (Fig. 3b). The spin accumulation, $\Delta\mu$, scales linearly with ΔT and changes sign when the hot and cold sides of the junction are interchanged. The Seebeck spin tunnelling coefficient can be defined uniquely as $S_{\text{st}} = \Delta\mu/\Delta T$ in units of volts per kelvin. It depends on properties of the non-magnetic electrode, in particular the spin relaxation time (Supplementary Information).

The sign and magnitude of $\Delta\mu$ depend crucially on the form of TSP as a function of energy, as shown in Fig. 3b for three functional forms selected for illustrative purposes. When the TSP is constant below E_{F} but decays above E_{F} (blue profile and symbols), the spin accumulation is positive for $T_{\text{Si}} > T_{\text{FM}}$, corresponding to majority spin accumulation with an induced magnetic moment in the silicon parallel to the

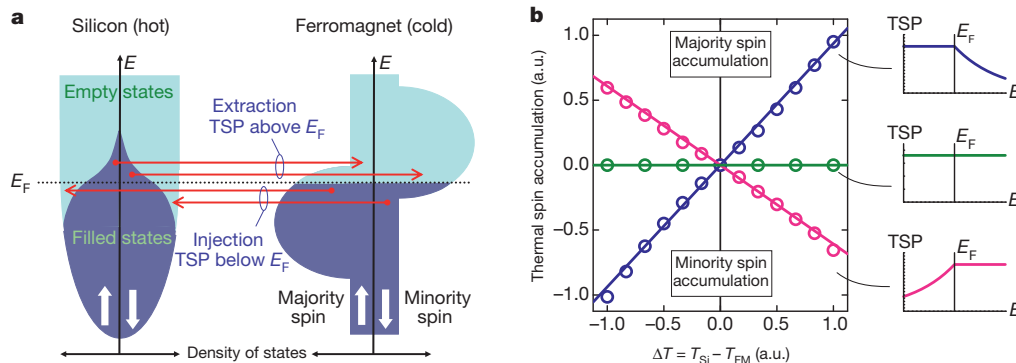


Figure 3 | Origin of Seebeck spin tunnelling and model calculation of salient characteristics. **a**, Spin-dependent density of states and its occupation for a tunnel contact with a hot silicon electrode and a cold ferromagnetic electrode, and the electron flows (four red arrows) that are induced by the temperature difference. For the cold ferromagnet, taken to be at $T_{\text{FM}} = 0$, all states below the Fermi energy, E_{F} , are filled (dark blue) and those above E_{F} are empty (light blue). For the hot silicon, a finite number of electrons with excess energy above E_{F} exist, together with a fraction of empty states below E_{F} . The sum of the four tunnel currents, two for each spin, results in a net flow of spin

magnetization of the ferromagnet. If the TSP is instead taken to decay with energy below E_{F} , the sign of $\Delta\mu$ is reversed (pink profile and symbols). Most strikingly, $\Delta\mu$ vanishes for any value of ΔT when the TSP is taken to be independent of energy but still non-zero (green profile and symbols). Hence, the induced spin accumulation does not depend on the sign of the TSP itself, but on the energy derivative of the TSP.

The sign of the experimental result (Fig. 2; accumulation of majority spins for silicon heating) is consistent with the prediction of the model calculation (Fig. 3b, blue curve; $\Delta\mu > 0$ for $\Delta T > 0$), considering that the TSPs for interfaces of transition-metal ferromagnets and Al_2O_3 in metal-based tunnel junctions are reported^{19,20} to decay significantly for energies above E_{F} . For the magnitude, the model calculation predicts a saturation of $\Delta\mu$ at about $10 \mu\text{V K}^{-1}$, and $T_{\text{Si}} - T_{\text{FM}}$ is estimated to be in the range of 0.1 K to a few kelvin at the most (Supplementary Information). The predicted $\Delta\mu$ is below about $10 \mu\text{V}$, which is much smaller than the measured signal. Thus, although the model provides a qualitative description of the salient observations, it does not provide quantitative agreement with the data. Part of this may be because it is a free-electron model, because the precise profile of the TSP as a function of energy is not known and because spin (and heat) transfer by inelastic, magnon-assisted tunnelling is not included. Such tunnelling has been shown to be important for the charge thermopower in magnetic tunnel junctions^{21–23}. These aspects may cause the model to underestimate the spin current across the tunnel barrier. However, the model also underestimates the electrically induced spin accumulation for which the spin current is rather well known; and control experiments¹⁶ and optical detection in spin-based light-emitting diodes²⁴ have clearly established that the large spin accumulation exists in the bulk bands. This suggests that the quantitative disagreement is at least in part related to the description of the conversion of injected spin current into a spin accumulation and, subsequently, into a detected voltage.

Next we demonstrate that the sign of the thermal spin current is reversed when ΔT is reversed. We compare devices in which the heating current is in the silicon (Fig. 4a) with devices in which it is in the ferromagnetic metal electrode (Fig. 4e). In the second of these cases, to improve the heating the metal electrode in the active part of the device consists of $\text{Ni}_{80}\text{Fe}_{20}$ (10 nm) and a gold layer only 10 nm thick, whereas in the first case the gold layer is 120 nm thick. A complication is that the current in the ferromagnet is accompanied by a voltage drop across it, such that any conventional magnetoresistance of the ferromagnet affects the signal via the term V_0 . For magnetic fields along the z axis, this results in a distortion of the Hanle curve that prevents accurate

determination of the induced spin accumulation. It was recently shown²⁵, however, that the spin accumulation can also be detected with an in-plane field parallel to the tunnel interface (in the x direction; see Fig. 2a), giving rise to a so-called inverted Hanle effect.

First we use the inverted Hanle effect to detect the spin accumulation induced by heating of the silicon, for the junction shown in Fig. 2. A clear, thermally induced inverted Hanle signal is observed (Fig. 4b), with sign inverted relative to the thermally induced Hanle effect (Fig. 2). Notably, the quadratic and linear scaling with, respectively, heating current and power (Fig. 4c, d) establishes the inverted Hanle effect as a good probe of thermally induced spin accumulation. We then perform a similar set of experiments with the heating current through the ferromagnet (Fig. 4f–h). A significant inverted Hanle signal is observed for heating current through the ferromagnet in either direction (Fig. 4f). The signal also scales quadratically with heating current and linearly with power (Fig. 4g, h), which is characteristic of a thermally induced spin accumulation. Most importantly, the sign of ΔV_{TH} is reversed; ΔV_{TH} is positive for $T_{\text{FM}} > T_{\text{Si}}$ and is negative for $T_{\text{Si}} > T_{\text{FM}}$. The reversal is not due to an unintentional difference between the tunnel contacts of the two devices, as electrical spin injection was similar for the two junctions (Supplementary information). Thus, the results establish another key feature of SST, namely that the thermal spin accumulation is reversed when ΔT is reversed. We note that the sharp features around zero field, which change in sign when the direction of the heating current is changed, arise from the aforementioned magnetoresistance of the ferromagnet. For a magnetic field applied along the x direction, these features can be clearly distinguished from the inverted Hanle curve and do not interfere with accurate determination of ΔV_{TH} (Supplementary Information). We also note that the true magnitude of the spin accumulation is the sum of the Hanle and inverted Hanle amplitudes²⁵.

We expect that the efficiency of Seebeck spin tunnelling can be greatly improved by better thermal design of our devices, which is far from optimal. The temperature difference across the thin tunnel barrier can also be optimized by using dielectrics with small interface thermal conductivity and by making use of phonon mismatch at interfaces. The mechanism of SST suggests the development of a new family of spin-tunnelling materials that, for instance, are designed to have strong variations in TSP around the Fermi energy. Conversely, SST can be used as a new probe of the energy dependence of the TSP. Thermal transfer of spin offers ways to design spintronic devices with fundamentally different characteristics. The fact that SST operates without tunnel current or voltage alleviates issues related to stability and breakdown of the tunnel

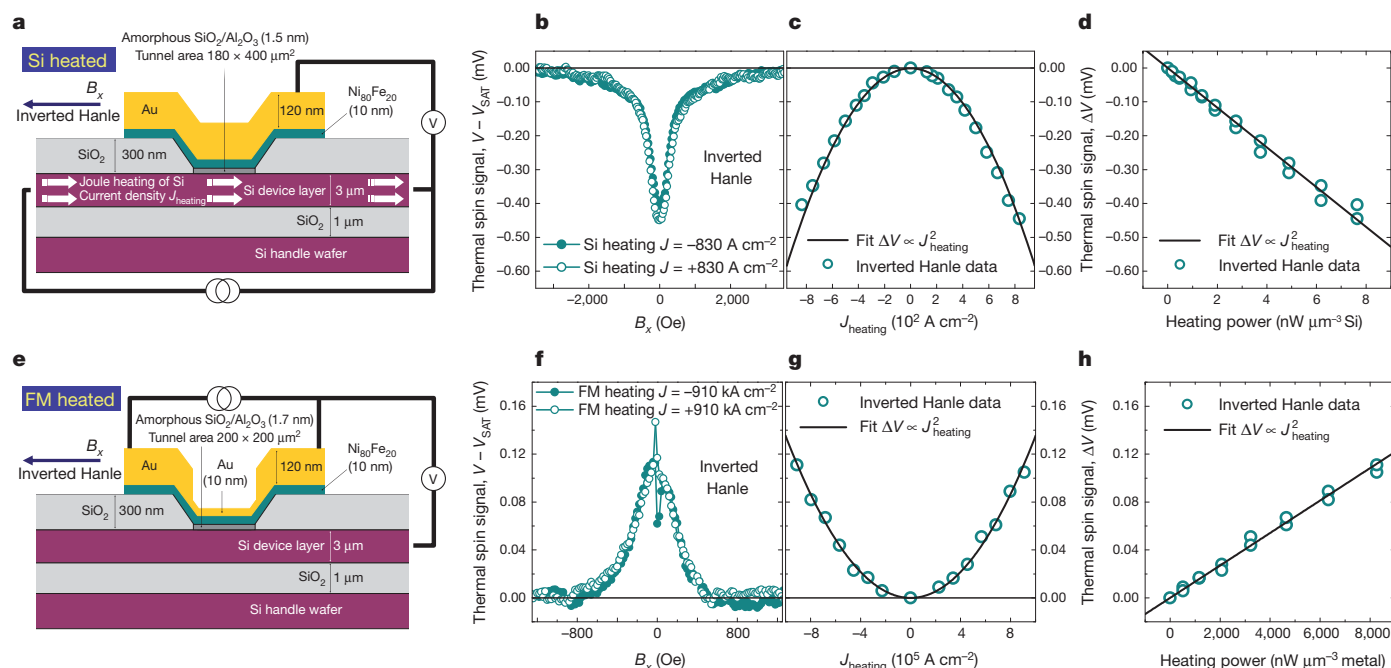


Figure 4 | Sign reversal of thermal spin current by heating the silicon or the ferromagnet. **a**, Layout of the device for silicon heating. **b**, Thermally induced spin accumulation for silicon heating in the same junction as in Fig. 2, but now detected with an external magnetic field, B_x , applied parallel to the tunnel interface (inverted Hanle geometry). **c**, **d**, Measured ΔV_{TH} (symbols) versus heater current density (**c**) and heating power (**d**), with corresponding fits. **e**, Layout of the device for heating of the ferromagnetic electrode (FM). In the

tunnel area, the metal electrode consists of 10-nm-thick $\text{Ni}_{80}\text{Fe}_{20}$ and 10-nm-thick gold. **f–h**, Similar sets of inverted Hanle data as in **b–d**, but now with the ferromagnetic electrode heated. For all data, the inverted Hanle signal is defined as the voltage at a given B_x minus the saturation voltage, V_{SAT} , at large B_x with voltage polarity consistently defined as $V_{Si} - V_{FM}$. All data were measured for p-type silicon with a resistivity of 11 mΩ cm at 300 K, for a tunnel contact area of $200 \mu\text{m} \times 200 \mu\text{m}$ and for a base temperature of 300 K.

barrier and decay of TSP at high bias that degrades electrical transfer of spin. It can also be linked to optics using lasers as the heating source, allowing optical creation and control of spin polarization in any semiconductor, including silicon, without the need for circularly polarized light and optical orientation. But above all, Seebeck spin tunnelling offers prospects to create spin currents in a more power-efficient way, either by itself or acting jointly with electrical spin injection, and allows the functional use of heat in electronic devices and circuits.

Received 23 February; accepted 24 May 2011.

Published online 29 June 2011.

- Uchida, K. *et al.* Observation of the spin Seebeck effect. *Nature* **455**, 778–781 (2008).
- Slachter, A., Bakker, F. L., Adam, J.-P. & van Wees, B. J. Thermally driven spin injection from a ferromagnet into a non-magnetic metal. *Nature Phys.* **6**, 879–882 (2010).
- Uchida, K. *et al.* Spin Seebeck insulator. *Nature Mater.* **9**, 894–897 (2010).
- Jaworski, C. M. *et al.* Observation of the spin-Seebeck effect in a ferromagnetic semiconductor. *Nature Mater.* **9**, 898–903 (2010).
- Žutić, I., Fabian, J. & Das Sarma, S. Spintronics: fundamentals and applications. *Rev. Mod. Phys.* **76**, 323–410 (2004).
- Chappert, C., Fert, A. & Nguyen van Dau, F. The emergence of spin electronics in data storage. *Nature Mater.* **6**, 813–823 (2007).
- Johnson, M. & Silsbee, R. H. Thermodynamic analysis of interfacial transport and of the thermomagnetolectric system. *Phys. Rev. B* **35**, 4959–4972 (1987).
- Fukushima, A. *et al.* Peltier effect in sub-micron-size metallic junctions. *Jpn. J. Appl. Phys.* **44**, L12–L14 (2005).
- Gravier, L., Serrano-Guisan, S., Reuse, F. & Ansermet, J.-Ph. Spin-dependent Peltier effect of perpendicular currents in multilayered nanowires. *Phys. Rev. B* **73**, 052410 (2006).
- Hatami, M., Bauer, G. E. W., Zhang, Q. & Kelly, P. J. Thermoelectric effects in magnetic nanostructures. *Phys. Rev. B* **79**, 174426 (2009).
- Slonczewski, J. C. Initiation of spin-transfer torque by thermal transport from magnons. *Phys. Rev. B* **82**, 054403 (2010).
- Bauer, G. E. W., MacDonald, A. H. & Maekawa, S. Spin caloritronics. *Solid State Commun.* **150**, 459–460 (2010).
- Xiao, J., Bauer, G. E. W., Uchida, K., Saitoh, E. & Maekawa, S. Theory of magnon-driven spin Seebeck effect. *Phys. Rev. B* **81**, 214418 (2010).
- Sinova, J. Spin Seebeck effect: thinks globally but acts locally. *Nature Mater.* **9**, 880–881 (2010).

- Adachi, H. *et al.* Gigantic enhancement of spin Seebeck effect by phonon drag. *Appl. Phys. Lett.* **97**, 252506 (2010).
- Dash, S. P., Sharma, S., Patel, R. S., de Jong, M. P. & Jansen, R. Electrical creation of spin polarization in silicon at room temperature. *Nature* **462**, 491–494 (2009).
- Min, B. C., Motohashi, K., Lodder, J. C. & Jansen, R. Tunable spin-tunnel contacts to silicon using low-work-function ferromagnets. *Nature Mater.* **5**, 817–822 (2006).
- Nolas, G. S., Sharp, J. & Goldsmid, H. J. *Thermoelectrics: Basic Principles and New Materials Developments* Ch. 1 (Springer, 2001).
- Valenzuela, S. O., Monsma, D. J., Marcus, C. M., Narayanamurti, V. & Tinkham, M. Spin polarized tunneling at finite bias. *Phys. Rev. Lett.* **94**, 196601 (2005).
- Park, B. G., Banerjee, T., Lodder, J. C. & Jansen, R. Tunnel spin polarization versus energy for clean and doped Al_2O_3 barriers. *Phys. Rev. Lett.* **99**, 217206 (2007).
- Wang, Z.-C., Su, G. & Gao, S. Spin-dependent thermal and electrical transport in a spin-valve system. *Phys. Rev. B* **63**, 224419 (2001).
- McCann, E. & Fal'ko, V. I. Giant magnetothermopower of magnon-assisted transport in ferromagnetic tunnel junctions. *Phys. Rev. B* **66**, 134424 (2002).
- McCann, E. & Fal'ko, V. I. A tunnel junction between a ferromagnet and a normal metal: magnon-assisted contribution to thermopower and conductance. *J. Magn. Magn. Mater.* **268**, 123–131 (2004).
- Jansen, R. *et al.* Electrical spin injection into moderately doped silicon enabled by tailored interfaces. *Phys. Rev. B* **82**, 241305 (2010).
- Dash, S. P. *et al.* Spin precession and decoherence near an interface with a ferromagnet. Preprint at (<http://arxiv.org/abs/1101.1691>) (2011).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to S. P. Dash for help with the device fabrication and discussions, T. Yoroza for the finite-element calculations and A. Yamamoto for making the finite-element program available to us. This work was financially supported by the program “Controlling spin dynamics in magnetic nanostructures” of the Netherlands Foundation for Fundamental Research on Matter (FOM).

Author Contributions J.-C.L.B. and R.J. designed the experiments. J.-C.L.B. and S.S. fabricated the devices. J.-C.L.B., S.S., H.S. and R.J. contributed to the measurements. R.J. developed the model calculation. All authors contributed to the planning, discussion and analysis of the research, and to the writing of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to R.J. (ron.jansen@aist.go.jp).

Tracking apex marine predator movements in a dynamic ocean

B. A. Block¹, I. D. Jonsen², S. J. Jorgensen¹, A. J. Winship², S. A. Shaffer³, S. J. Bograd⁴, E. L. Hazen⁴, D. G. Foley⁴, G. A. Breed^{2,5}, A.-L. Harrison⁵, J. E. Ganong¹, A. Swithenbank¹, M. Castleton¹, H. Dewar⁶, B. R. Mate⁷, G. L. Shillinger¹, K. M. Schaefer⁸, S. R. Benson⁹, M. J. Weise⁵, R. W. Henry⁵ & D. P. Costa⁵

Pelagic marine predators face unprecedented challenges and uncertain futures. Overexploitation and climate variability impact the abundance and distribution of top predators in ocean ecosystems^{1–4}. Improved understanding of ecological patterns, evolutionary constraints and ecosystem function is critical for preventing extinctions, loss of biodiversity and disruption of ecosystem services. Recent advances in electronic tagging techniques have provided the capacity to observe the movements and long-distance migrations of animals in relation to ocean processes across a range of ecological scales^{5,6}. Tagging of Pacific Predators, a field programme of the Census of Marine Life, deployed 4,306 tags on 23 species in the North Pacific Ocean, resulting in a tracking data set of unprecedented scale and species diversity that covers 265,386 tracking days from 2000 to 2009. Here we report migration pathways, link ocean features to multispecies hotspots and illustrate niche partitioning within and among congener guilds. Our results indicate that the California Current large marine ecosystem and the North Pacific transition zone attract and retain a diverse assemblage of marine vertebrates. Within the California Current large marine ecosystem, several predator guilds seasonally undertake north–south migrations that may be driven by oceanic processes, species-specific thermal tolerances and shifts in prey distributions. We identify critical habitats across multinational boundaries and show that top predators exploit their environment in predictable ways, providing the foundation for spatial management of large marine ecosystems.

Large marine predators are important top consumers in marine ecosystems, and their depletion can have cascading effects on lower trophic levels in both coastal and open-ocean waters^{3,7,8}. Although some taxa are protected internationally from direct harvest, intense over-exploitation by high-seas fisheries and anthropogenic environmental change are threatening global marine biodiversity. The management and conservation of highly migratory marine predators depends on understanding how movements relate to ocean processes; however, the dynamic relationships among behaviour, physiology and oceanography over a range of temporal and spatial scales remains unclear^{4,9,10}.

Tagging of Pacific Predators (TOPP) examined the distribution of species that inhabit and migrate into the coastal upwelling region off the west coast of North America. Using simultaneous tag deployments, we tracked seven top-predator guilds (sharks, tunas, albatrosses, shearwaters, turtles, pinnipeds and rorqual whales) concurrently for nearly a decade. Habitat use, residency patterns and migration pathways to North American waters were previously unknown or poorly documented for most of these animals^{5,6,11–13}.

Deployments of 4,306 electronic tags yielded 1,791 individual animal tracks from 23 species, totalling 265,386 animal tracking days (Fig. 1 and Supplementary Tables 1–3). The tracking technologies used had different location precisions, and include ARGOS satellite tags

($n = 1,183$), archival and satellite transmitting tags ($n = 1,008$) and archival geolocation tags ($n = 2,115$); the latter two provide estimates of position based on sunrise and sunset times and sea surface temperature (SST). We used a state-space modelling approach¹⁴ to estimate daily locations, and their uncertainty, from each track while accounting for the differing precisions of the tracking technologies.

The tracking data show that the California Current large marine ecosystem (CCLME; Supplementary Fig. 1) is an important habitat (Figs 2–4) for tunas (Pacific bluefin, *Thunnus orientalis*; yellowfin, *Thunnus albacares*; albacore, *Thunnus alalunga*), sharks (shortfin mako, *Isurus oxyrinchus*; white, *Carcharodon carcharias*; salmon, *Lamna ditropis*; blue, *Prionace glauca*; common thresher, *Alopias vulpinus*), cetaceans (blue whale, *Balaenoptera musculus*), pinnipeds (northern elephant seal, *Mirounga angustirostris*; California sea lion, *Zalophus californianus*), seabirds (Laysan albatross, *Phoebastria immutabilis*; black-footed albatross, *Phoebastria nigripes*; sooty shearwater, *Puffinus griseus*) and sea turtles (leatherback, *Dermochelys coriacea*; loggerhead, *Caretta caretta*). Annual migratory periodicity was evident in the movements of many tagged animals that showed fidelity to the cool, nutrient-rich waters of the CCLME (Figs 2 and 3). Extended residency within the CCLME was revealed by examining tracks that spanned multiple seasons, using a behavioural switching state-space model¹⁴. Numerous species (leatherback sea turtles, black-footed albatrosses, sooty shearwaters, bluefin tunas and salmon sharks; Fig. 2b) had a strong attraction to the CCLME and undertook long migrations (>2,000 km) from the western, central or south Pacific basin. Some species (bluefin and yellowfin tunas; mako, white and salmon sharks; blue whales; male elephant seals; and leatherback sea turtles; Fig. 3a and Supplementary Fig. 2) undertook a seasonally recurring north–south migration in the North Pacific and within the CCLME. Other taxa undertook movements between near-shore and offshore waters, with a residency period within the CCLME or the Gulf of Alaska, followed by an offshore migration that ranged into the North Pacific transition zone (NPTZ) (female elephant seals, salmon sharks and Laysan albatrosses), the subtropical gyre and north equatorial current (blue and mako sharks and leatherback sea turtles), or the ‘Café’ region of the eastern Pacific and the Hawaiian Islands (white sharks; albacore tunas, *Thunnus alalunga*; and black-footed albatrosses). The mechanisms and cues underlying fidelity to seasonally modulated migration pathways are not entirely known, but may represent a capacity to discriminate among areas of seasonal significance for foraging or reproduction.

The quarterly predator density in relation to the median satellite-derived SST measurements within the CCLME (Fig. 3b and Supplementary Fig. 3) suggests that seasonal warming (quarter three) along the Baja California peninsula may trigger northward movements of predator populations. Increased metabolic expenditures in ectotherms

¹Stanford University, Biology Department, Hopkins Marine Station, Pacific Grove, California 93950, USA. ²Dalhousie University, Department of Biology, Halifax, Nova Scotia B3H 4J1, Canada. ³San Jose State University, Department of Biological Sciences, San Jose, California 95192, USA. ⁴NOAA Southwest Fisheries Science Center, Environmental Research Division, Pacific Grove, California 93950, USA. ⁵University of California, Santa Cruz, Department of Ecology & Evolutionary Biology, Long Marine Laboratory, Santa Cruz, California 95060, USA. ⁶NOAA Southwest Fisheries Science Center, Fisheries Research Division, La Jolla, California 92037, USA. ⁷Oregon State University, Hatfield Marine Science Center, Newport, Oregon 97365, USA. ⁸Inter-American Tropical Tuna Commission, La Jolla, California 92037, USA. ⁹NOAA Southwest Fisheries Science Center, Protected Resources Division, La Jolla, California 92037, USA.

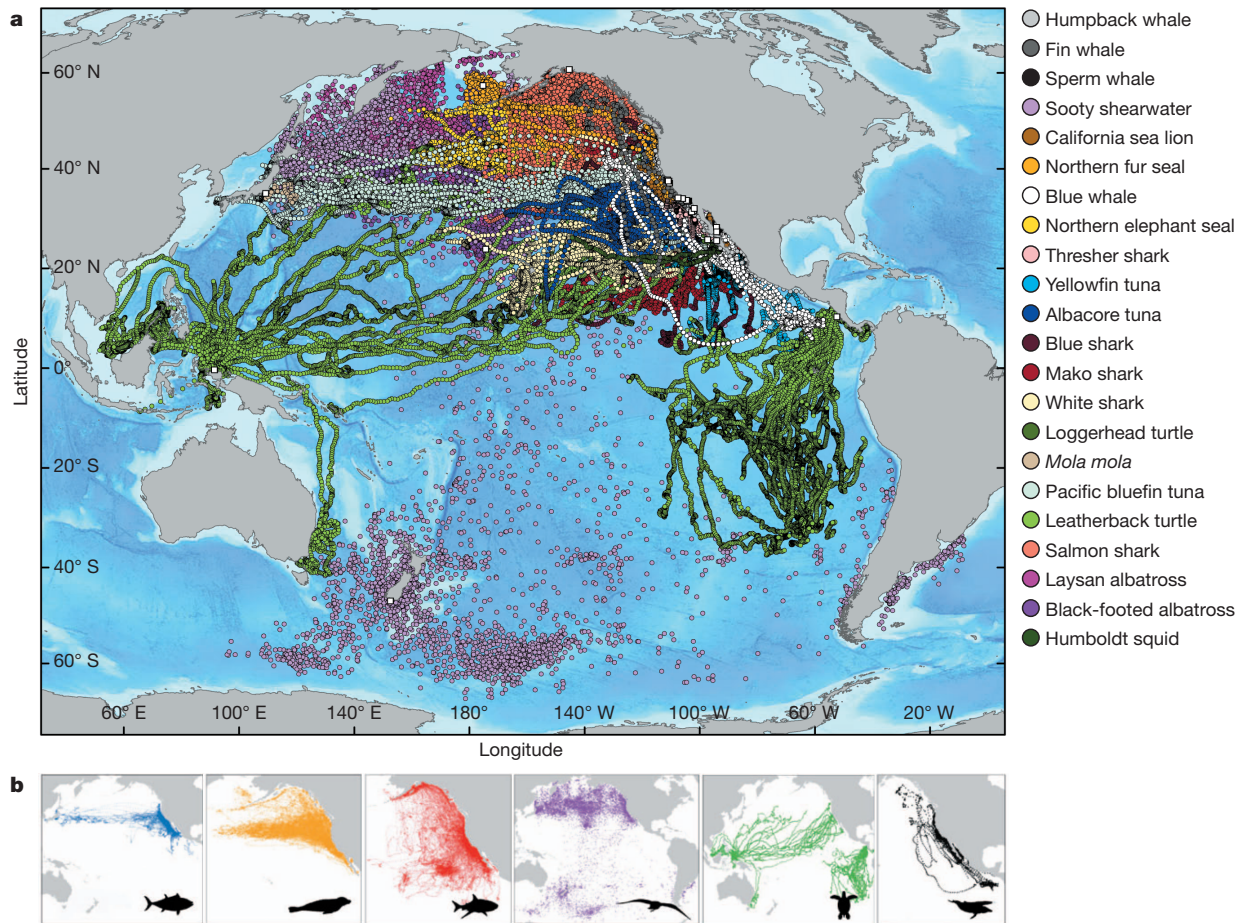


Figure 1 | All TOPP species state space position estimates and distribution from electronic tagging. **a**, Daily mean position estimates (circles) and annual median deployment locations (white squares) of all tagged species. **b**, Daily mean position estimates of the major TOPP guilds (from left): tunas (yellowfin, bluefin and albacore), pinnipeds (northern elephant seals, California sea lions

as the ocean warms, subsurface hypoxia¹⁵ and declines in primary productivity (Fig. 3b and Supplementary Fig. 3g) all potentially decrease the suitability of this habitat. In the northern waters of the CCLME (Fig. 3b and Supplementary Fig. 3a–d), cooler temperatures (9–11 °C), particularly during the spring upwelling season, limit cardiac function^{13,16} in many fishes and sharks, reducing their capacity to exploit these environs. Predators were observed moving south into warmer water despite regions of persistent productivity in the north (Fig. 3b). A similar relationship may hold for endotherms, such as seabirds, pinnipeds and whales, but the linkage is indirect as the physiological constraints are largely on their prey.

and northern fur seals), sharks (salmon, white, blue, common thresher and mako), seabirds (Laysan and black-footed albatrosses and sooty shearwaters), sea turtles (leatherback and loggerhead) and cetaceans (blue, fin, sperm and humpback whales).

Tracking data typically were acquired for less than one year; however, for some taxa (tunas, turtles and sharks) multiyear tracks were obtained (maximum of 1,380 days, for salmon sharks; Supplementary Table 1 and Supplementary Fig. 4). To account for abbreviated tracks and for variation in sample size among taxa, we applied time weighting and species normalization schemes before examining multispecies density patterns (Supplementary Methods). The relative spatial density of large marine predators indicates that high-use areas in the northeastern Pacific occur both on the high seas (NPTZ; Fig. 4a and Supplementary Fig. 1) and within the exclusive economic zones of North America (CCLME; Fig. 4b). These observed patterns differ markedly from those expected

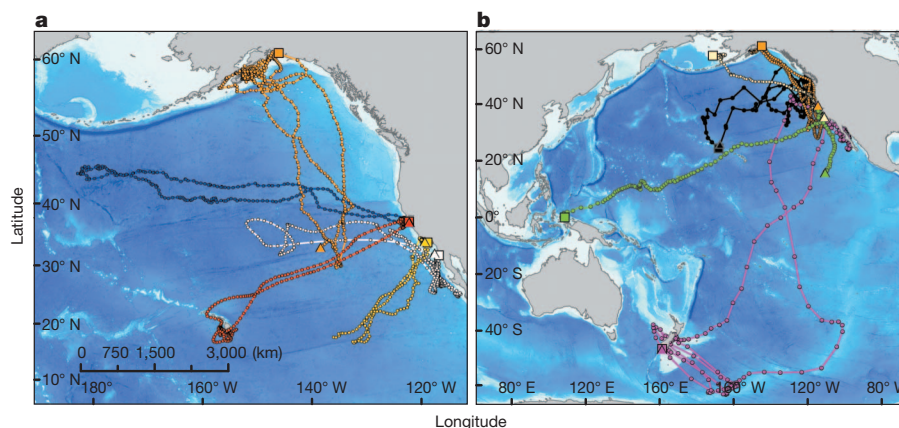


Figure 2 | Fidelity and attraction to the CCLME. **a**, Examples of pelagic predators released and electronically tracked in the CCLME that show fidelity to deployment locations and the CCLME. We show the release locations (square), pop-up satellite end point locations (triangle) and daily mean positions (circles) of the following species: yellowfin tuna (yellow), bluefin tuna (white), white shark (red), elephant seal (blue) and salmon shark (orange). **b**, Individual tracks of pelagic animals released >2,000 km away from the CCLME that are indicative of cross-basin or ecosystem attraction to, and temporary residency within, the eastern North Pacific. Symbols are as in **a**, for leatherback sea turtles (green), sooty shearwaters (pink), fur seals (pale yellow), black-footed albatrosses (black) and salmon sharks (orange).

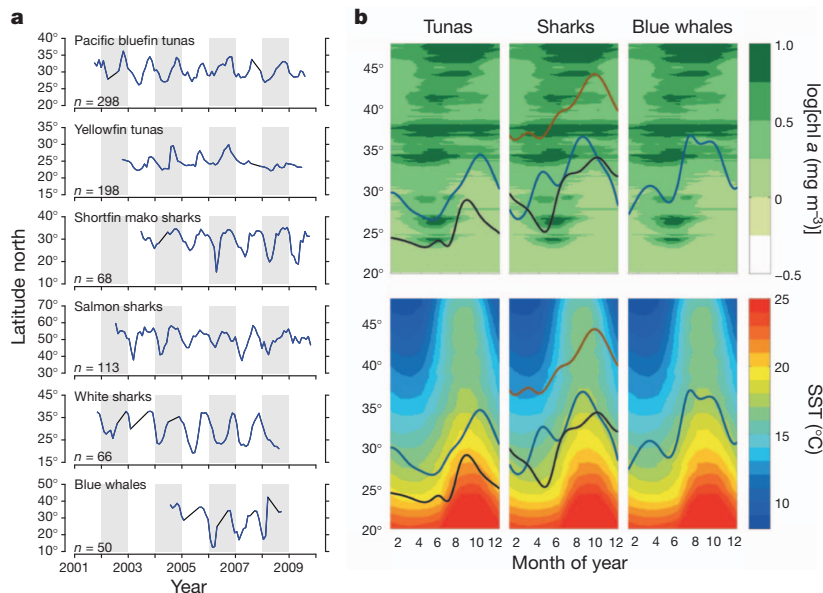


Figure 3 | Latitudinal migration cycles and seasonal climatologies within the CCLME.

a, Monthly mean latitudes of predators residing within or migrating to the CCLME. Black line segments denote gaps where no data were available. Sample sizes indicate the numbers of individual tracks contributing to the time series. **b**, Seasonal climatologies in the California Current for tunas (Pacific bluefin, blue; yellowfin, black), sharks (salmon shark, brown; shortfin mako, black; white shark, blue) and blue whales relative to median chlorophyll *a* densities and SST values between 2000 and 2009.

under a null model of random diffusion from tag deployment locations (Supplementary Fig. 5 and Supplementary Methods), confirming that the observed density patterns are not driven by the tag deployment locations.

The CCLME is a highly retentive area for many species tagged there, and is an attractive area for animals undergoing long migrations from the western and central North Pacific and the Gulf of Alaska (Figs 2 and 4a, b). Pacific bluefin and yellowfin tunas spent significantly more time in the CCLME than expected on the basis of null model simulations (Supplementary Fig. 6). Several species (tunas, white, mako and salmon sharks, leatherback sea turtles and blue whales) had substantial residency periods within, or were return migrants to, the CCLME, according to behavioural inferences from a switching state-space model¹⁴ (Fig. 4c). Additionally, all species tagged outside the CCLME spent significantly more time on average in the CCLME than expected on the basis of null model simulations (Supplementary Fig. 6). Representatives from several guilds exhibited cross-basin migrations (>2,000 km) into the CCLME from the western (leatherback sea turtles and bluefin tunas), central (black-footed albatrosses and salmon sharks) and south Pacific basins (sooty shearwaters; Fig. 2b). The retention within and attraction to the CCLME is consistent with the high productivity of this region, which supports large biomasses of krill, sardines, anchovies, salmon, groundfish and squid that provide a predictable forage base for top predators^{2,17}. The NPTZ is another important region, serving as an east–west migration corridor (Fig. 4a) and foraging region for elephant seals, salmon and blue sharks (Fig. 4c), Laysan and black-footed albatrosses, and bluefin tunas (Fig. 1). This is a complex region encompassing an abrupt north-to-south transition between sub-Arctic and subtropical water masses with dynamic frontal regions¹⁸.

To investigate which aspects of the biophysical environment putatively attract these predators, we explored both presence/absence and relative habitat use with generalized additive mixed models¹⁹. We examined the collective response of 16 marine predator species to environmental covariates (Supplementary Fig. 7 and Supplementary Methods). In the binary presence/absence model, predator incidence showed a strong positive relationship with SST across a broad temperature range that peaked near 15 °C (Supplementary Fig. 7a). Tagged animals occupied a small portion of cool, nutrient-rich water in coastal regions and northern latitudes in preference to the broadly available warm oligotrophic waters in lower latitudes (Supplementary Fig. 3). Relative density models showed a strong positive relationship between density and chlorophyll *a* suggesting that the suite of tagged species preferentially occupy regions of high productivity (Supplementary Fig. 7c). The observed patterns of predator distribution in this study may be indicative of trade-offs between access to areas of higher productivity

and physiological thermal tolerances, either directly by the predators or indirectly by their lower-trophic-level prey.

To examine how closely related taxa partition marine resources, we compared thermal preferences from *in situ*, tag-based SST measurements for sympatrically occurring species within three guilds (albatrosses, tunas and sharks). Differences in habitat use evident among congeneric species illustrate how more-recently divergent species partition the oceanic environment (Fig. 5). During the June–November post-breeding phase, black-footed albatrosses were associated with a broader range of higher SSTs primarily in the eastern Pacific, whereas Laysan albatrosses were associated with a narrower range of lower SSTs in the western and central North Pacific (Fig. 5a and Supplementary Table 9). Bluefin tunas ranged farther north in the colder waters of the CCLME, whereas yellowfin tunas occupied warmer waters in the southern CCLME (Fig. 5b and Supplementary Table 10). These differences are consistent with physiological specializations in bluefin tuna cardiac performance¹⁶. The lamnid sharks had a more complex separation of habitats. Salmon sharks, with their cold-tolerant cardiac physiology¹³, occupied the cooler, sub-Arctic waters in the North Pacific. Most salmon sharks, but not all, migrated seasonally into the warmer NPTZ and CCLME waters (Fig. 5c and Supplementary Table 11). White sharks overlapped with salmon sharks in the near-shore CCLME but also migrated into warmer, offshore waters of the subtropical gyre¹² and the Hawaiian Islands (Fig. 5c and Supplementary Table 11). Shortfin mako sharks were distributed throughout the CCLME and into the subtropical gyre but occupied a thermal range intermediate between the two modes of the white shark range (Fig. 5c and Supplementary Table 11).

Species such as black-footed albatrosses, sooty shearwaters, leatherback sea turtles and salmon sharks have evolved migration or foraging strategies that entail long-distance journeys to productive foraging habitats within the CCLME and the NPTZ. Shearwaters⁶ and leatherback sea turtles tagged in the south and western Pacific arrived in the central and northern portions of the CCLME each year, when SSTs were seasonally highest. These patterns indicate that resources within the CCLME and the NPTZ are valuable to species throughout the Pacific basin, highlighting the importance of maintaining ecological connectivity across hemispheres. It is likely that other areas of the Pacific, not extensively covered by our tagging efforts, are similarly important. For example, estimates of global marine species richness peak around Southeast Asia, and richness of primarily oceanic taxa peaks at latitude 30° (north or south) along boundary currents such as the Kuroshio in the western Pacific²⁰.

The concentration and residency of multiple predators at predictable locations in the Pacific reveals the importance of these regions as biological hotspots. Given the high density of tracked animals in these regions, predators may have an important top-down role in structuring

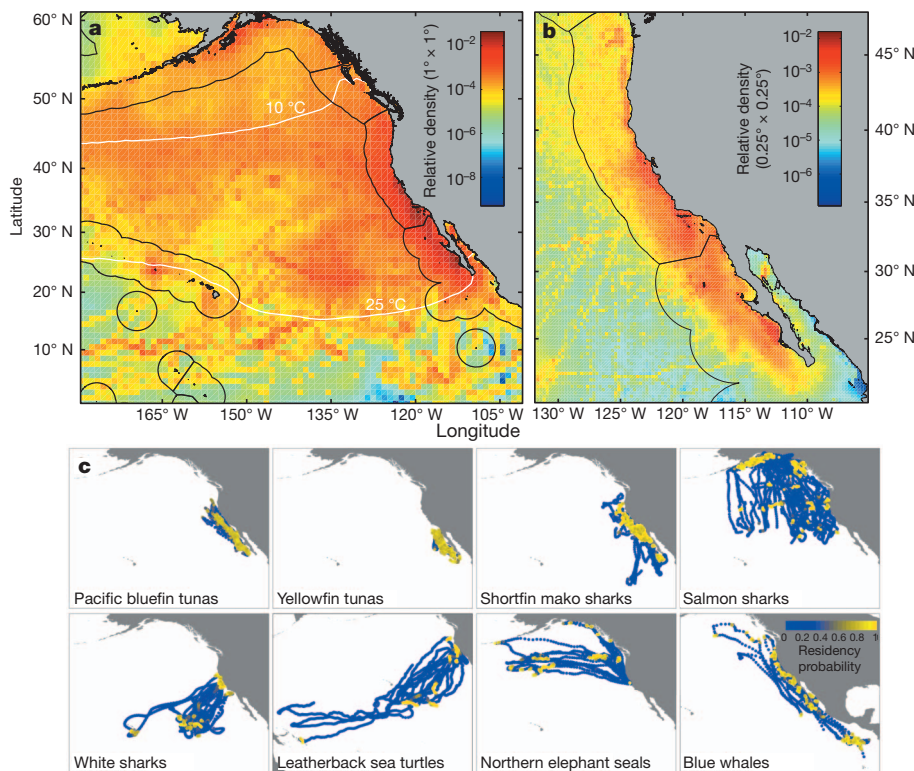


Figure 4 | Predator density maps and residency patterns. **a**, Density of large marine predators within the eastern North Pacific. Densities of the time-weighted and species-normalized position estimates of all tagged individuals were summed within $1^\circ \times 1^\circ$ grid cells. **b**, Density of large marine predators within the CCLME at a $0.25^\circ \times 0.25^\circ$ resolution. **c**, Patterns of resident (slow, area-restricted movements) versus transient (fast, directed movements) behaviours of predators that primarily occupied or migrated to the CCLME, estimated using a switching state-space model. The coloured points grading from blue to yellow display the posterior mean probability of the resident behaviour associated with each daily mean position estimate. Each panel displays residency patterns for ten individuals. Uncertainty in position estimates in **a** and **b** is included by calculating densities using all 2000 Markov chain Monte Carlo samples from the joint posterior distributions of the daily positions, rather than using only the posterior means. SST contours in **a** are denoted by solid white lines. Exclusive economic zones are delineated by solid black lines.

these ecosystems. The impact that the removal of top predators has had on the trophic dynamics of coastal and oceanic ecosystems is contentious^{3,21}. For example, a failure to manage marine resources effectively in the Atlantic has led to a decline of pelagic species such as the Atlantic bluefin tuna³ and the porbeagle shark²², and has brought demersal fish, such as cod, to the brink of ecological extinction²³. In the Pacific, commercial fisheries removed over 50 million tonnes of tunas and sharks from pelagic ecosystems between 1950 and 2004²¹, resulting in significant declines in populations of large fish. Furthermore, unsustainable levels of fishery bycatch mortality has decimated populations of Pacific sea turtle (leatherback and loggerhead) and threatened some populations of albatross^{24,25}. The impact of these top-predator biodiversity losses on ecosystem stability is unknown. Establishing the population baselines for top-predator populations and conserving remaining populations in regions such as the CCLME, and quantifying the ecosystem services they provide, remains an important management priority²⁶. Our study provides a clearer understanding of where predators are concentrated in the North Pacific and which environmental parameters can be used to predict their occurrence.

Tracking data can be used to examine behaviours and distributions, inform population assessments and improve estimates of natural and fishery-induced mortality^{27–29}. Despite their high monetary and conservation value, population assessments of many predators in this study are rare or non-existent. Long-term observations of predator movements provide information on the spatial extent of highly migratory populations and potential rates of exchange among them. High archival tag returns (Supplementary Table 1) for bluefin (56%) and yellowfin (39%) tunas indicate that high fishery-induced mortality currently occurs within the CCLME. The recovery of cetacean populations, such as blue whales, which are endangered, can be enhanced through the use of electronic tracking data to identify high-use areas and coordinate policy actions that mitigate anthropogenic risks³⁰ (for example ship strikes). Finally, an improved understanding of the horizontal and vertical spatiotemporal distribution of leatherback sea turtles and North Pacific albatrosses in relation to pelagic fisheries could better inform management protocols (for example time–area closures and gear mitigation) to reduce bycatch of these endangered and threatened species.

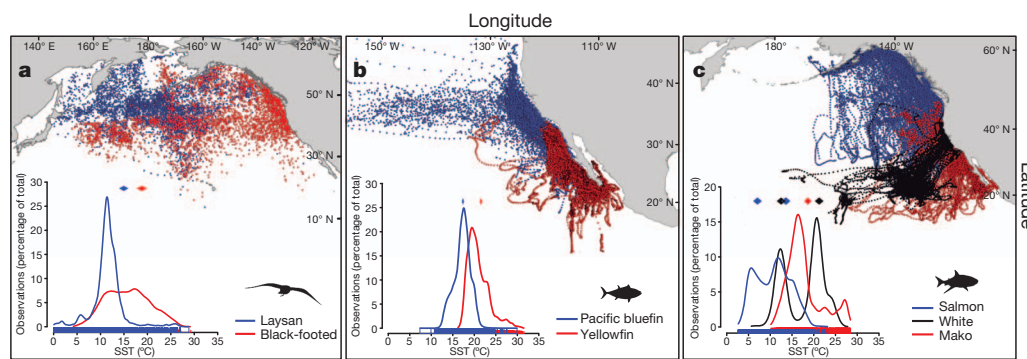


Figure 5 | Niche separation within three predator guilds. Spatial distribution and thermal habitat use (insets) across three guilds of sympatric species: Laysan and black-footed albatrosses ($n = 55$ individuals, 8,743 daily SSTs; **a**), Pacific bluefin and yellowfin tunas ($n = 376$ individuals, 75,177 daily SSTs; **b**), lamnid (salmon, mako and white) sharks ($n = 137$ individuals, 12,971 daily SSTs; **c**).

SST profiles are daily means of tag-derived SSTs. Linear mixed-effects model estimates of mean ($\pm 95\%$ confidence interval) SST for each species are displayed as diamonds at the top of each inset graph. The distribution of daily mean SSTs for each species is displayed along the x axes.

The decadal record of multitaxa electronic tagging generated by TOPP provides information on cross-border movements between US, Mexican and Canadian waters that has important implications for conservation and management efforts. This extensive data set serves as a baseline for monitoring and forecasting these seasonal patterns and assessing shifts in abundance among countries. Linking movements to ocean temperatures can help predict the potential ecosystem-level effects of climate variability. This baseline information could provide the foundation for launching a management scheme on a scale never before achieved for a pelagic ecosystem. International policy vehicles such as a UNESCO Marine World Heritage designation could provide an effective approach to generating global interest and, in turn, funding to support ecosystem-based management efforts within the CCLME. Furthermore, if a conservation corridor was extended across the NPTZ, key ecological foraging hotspots and migratory corridors that link the eastern and western Pacific basins for transoceanic migrants could be sustainably managed. Without an aggressive effort to zone and effectively manage these resources, the predator populations they support will decline and the biodiversity of this open-ocean wilderness will be irreplaceably lost.

METHODS SUMMARY

We used a Bayesian state-space model¹⁴ to account for observation error among tag types and to obtain regular daily location and uncertainty estimates. Argos satellite errors were estimated previously¹⁴. Errors for archival geolocation and pop-up satellite archival tags were estimated from taxon-specific double-tagging experiments (Supplementary Methods) and captive sea pen studies. We used a behavioural switching state-space model¹⁴ to infer patterns of residency versus transience from a subset of species' tracks that at least partly inhabited the CCLME.

Unnaturally abbreviated tracks were normalized by weighting each location estimate by the inverse of the number of individuals of that species that had location estimates for the same relative day of their track. We imposed a threshold relative day of track (85th percentile) above which locations received the same weight as on the threshold day. We normalized the data set by species to account for the over-representation of species with large sample sizes. The normalized and weighted tracks were summed within $1^\circ \times 1^\circ$ or $0.25^\circ \times 0.25^\circ$ grid cells to obtain relative densities of 16 predator species across the North Pacific and within the CCLME, respectively. The influence of the tag deployment locations on the spatial density pattern was assessed through null model simulations. The relative density of predator species in the $1^\circ \times 1^\circ$ grid cells was related to oceanographic variables using generalized additive mixed models¹⁹, with season and year as random covariates. Satellite-derived chlorophyll *a*, SST and sea surface height anomalies and variances, and wind stress curl, were summarized quarterly in each grid cell between June 2002 and June 2009. A two-model approach dealt with over-dispersion, non-normality and non-constant variance of the relative density data. We used a binary presence/absence habitat model with a logit link function to model the occurrence of predators within grid cells. All cells with non-zero density were log-transformed and fitted to environmental correlates to create a relative-use habitat model.

Further details of the methods are included in Supplementary Information.

Received 31 December 2010; accepted 1 April 2011.

Published online 22 June 2011.

- Robinson, R. A. *et al.* Travelling through a warming world: climate change and migratory species. *Endanger. Species Res.* **7**, 87–99 (2009).
- Chavez, F. P. *et al.* From anchovies to sardines and back: multidecadal change in the Pacific ocean. *Science* **299**, 217–221 (2003).
- Myers, R. A. & Worm, B. Rapid worldwide depletion of predatory fish communities. *Nature* **423**, 280–283 (2003).
- Costa, D. P. *et al.* Approaches to studying climate change and habitat selection of Antarctic pinnipeds. *Integr. Comp. Biol.* **50**, 1018–1030 (2010).
- Shillinger, G. *et al.* Persistent leatherback turtle migration corridor presents opportunities for conservation. *PLoS Biol.* **6**, e171 (2008).
- Shaffer, S. A. *et al.* Migratory shearwaters integrate oceanic resources across the Pacific Ocean in an endless summer. *Proc. Natl Acad. Sci. USA* **103**, 12799–12802 (2006).
- Scheffer, M., Carpenter, S. & de Young, B. Cascading effects of overfishing marine systems. *Trends Ecol. Evol.* **20**, 579–581 (2005).
- Springer, A. M. *et al.* Sequential megafaunal collapse in the North Pacific Ocean: an ongoing legacy of industrial whaling? *Proc. Natl Acad. Sci. USA* **100**, 12223–12228 (2003).
- Steele, J. H. & Henderson, E. W. Coupling between physical and biological scales. *Phil. Trans. R. Soc. Lond. B* **343**, 5–9 (1994).
- Whitehead, H., McGill, B. & Worm, B. Diversity of deep-water cetaceans in relation to temperature: implications for ocean warming. *Ecol. Lett.* **11**, 1198–1207 (2008).

- Bailey, H. *et al.* Behavioural estimation of blue whale movements in the Northeast Pacific from state-space model analysis of satellite tracks. *Endanger. Species Res.* **10**, 93–106 (2009).
- Jorgensen, S. J. *et al.* Philopatry and migration of Pacific white sharks. *Proc. R. Soc. Lond. B* **277**, 679–688 (2010).
- Weng, K. C. *et al.* Satellite tagging and cardiac physiology reveal niche expansion in salmon sharks. *Science* **310**, 104–106 (2005).
- Jonsen, I. D., Flemming, J. M. & Myers, R. A. Robust state-space modeling of animal movement data. *Ecology* **86**, 2874–2880 (2005).
- Prince, E. & Goodyear, C. Hypoxia-based habitat compression of tropical pelagic fishes. *Fish. Oceanogr.* **15**, 451–464 (2006).
- Shiels, H. A., Di Maio, A., Thompson, S. & Block, B. A. Warm fish with cold hearts: thermal plasticity of excitation-contraction coupling of bluefin tuna. *Proc. R. Soc. Lond. B* **278**, 18–27 (2011).
- PICES. *Marine Ecosystems of the North Pacific*. PICES Spec. Publ. 1 (North Pacific Marine Science Organization, 2004).
- Polovina, J. J., Howell, E., Kobayashi, D. R. & Seki, M. P. The transition zone chlorophyll front, a dynamic global feature defining migration and forage habitat for marine resources. *Prog. Oceanogr.* **49**, 469–483 (2001).
- Wood, S. *Generalized Additive Models: An Introduction with R* (Chapman & Hall/CRC, 2006).
- Tittensor, D. P. *et al.* Global patterns and predictors of marine biodiversity across taxa. *Nature* **466**, 1098–1101 (2010).
- Sibert, J., Hampton, J., Kleiber, P. & Maunder, M. Biomass, size and trophic status of top predators in the Pacific ocean. *Science* **314**, 1773–1776 (2006).
- Campana, S. E. *et al.* Population dynamics of the porbeagle in the northwest Atlantic Ocean. *N. Am. J. Fish. Manag.* **22**, 106–121 (2002).
- Swain, D. P. & Chouinard, G. A. Predicted extirpation of the dominant demersal fish in a large marine ecosystem: Atlantic cod (*Gadus morhua*) in the southern Gulf of St. Lawrence. *Can. J. Fish. Aquat. Sci.* **65**, 2315–2319 (2008).
- Peckham, S. H. *et al.* Small-scale fisheries bycatch jeopardizes endangered Pacific loggerhead turtles. *PLoS ONE* **2**, e1041 (2007).
- Lewison, R. L. & Crowder, L. B. Estimating fishery bycatch and effects on a vulnerable seabird population. *Ecol. Appl.* **13**, 743–753 (2003).
- Halpern, B. S. *et al.* Mapping cumulative human impacts to California Current marine ecosystems. *Conserv. Lett.* **2**, 138–148 (2009).
- Hays, G. C. *et al.* Satellite telemetry suggests high levels of fishing-induced mortality in marine turtles. *Mar. Ecol. Prog. Ser.* **262**, 305–309 (2003).
- Heupel, M. & Simpfendorfer, C. Estimation of mortality of juvenile blacktip sharks, *Carcharhinus limbatus*, within a nursery area using telemetry data. *Can. J. Fish. Aquat. Sci.* **59**, 624–632 (2002).
- Kurota, H. *et al.* A sequential Bayesian methodology to estimate movement and exploitation rates using electronic and conventional tag data: application to Atlantic bluefin tuna (*Thunnus thynnus*). *Can. J. Fish. Aquat. Sci.* **66**, 321–342 (2009).
- Berman-Kowalewski, M. *et al.* Association between blue whale (*Balaenoptera musculus*) mortality and ship strikes along the California coast. *Aquat. Mamm.* **36**, 59–66 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This manuscript is the culmination of a Census of Marine Life cross-project synthesis between TOPP and Future of Marine Animal Populations (FMAP). Funding for this work was provided by the Sloan Foundation's Census of Marine Life programme. TOPP research was funded by the Sloan, Packard and Moore foundations. FMAP was funded by the Sloan Foundation. Electronic tagging and tracking in TOPP was also supported by the Office of Naval Research, the NOAA, the E&P Sound and Marine Life JIP under contract from the OGP, and the Monterey Bay Aquarium Foundation. We thank the TOPP scientific teams and all those who supported animal tagging efforts, R. Kochevar and D. Kohrs for their dedication and their effort on behalf of the Census of Marine Life. We are grateful to the numerous captains and crews who provided ship time and logistical support, and to the US Fish and Wildlife Service in Hawaii for further logistical support. We thank the Mexican authorities and collaborating TOPP partners (O. Sosa-Nishizki) for permitting and assisting in research in their waters. All animal research was conducted in accordance with IACUC protocols from Stanford University and the University of California.

Author Contributions This synthesis study was initiated by B.A.B. and I.D.J. The TOPP project was designed and coordinated by B.A.B., D.P.C. and S.J.B. B.A.B., S.J.J., H.D. and K.M.S. designed experiments and deployed electronic tags on fish and sharks. D.P.C., S.A.S., R.W.H., M.J.W. and B.R.M. designed experiments and deployed electronic tags on marine mammals and birds. G.L.S., B.A.B. and S.R.B. designed experiments and deployed electronic tags on sea turtles. Tracking data were compiled by S.J.J., S.A.S., G.A.B., A.-L.H., B.A.B., G.L.S. and M.C. Data management was coordinated by A.S. and J.E.G. Oceanographic data were compiled by S.J.B., E.L.H. and D.G.F. Analyses were conducted by A.J.W., S.J.J., I.D.J., G.A.B., E.L.H., D.G.F., A.-L.H., J.E.G. and A.S. Figures were created by B.A.B., M.C., A.-L.H., I.D.J., S.J.J., A.J.W., J.E.G., A.S., E.L.H. and D.G.F. The manuscript was drafted by B.A.B. and edited by I.D.J., D.P.C., S.J.J., S.A.S., S.J.B., E.L.H., A.-L.H., A.J.W., H.D., G.L.S. and B.R.M.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to B.A.B. (bblock@stanford.edu).

NMDA receptor blockade at rest triggers rapid behavioural antidepressant responses

Anita E. Autry¹, Megumi Adachi¹, Elena Nosyreva², Elisa S. Na¹, Maarten F. Los¹, Peng-fei Cheng¹, Ege T. Kavalali² & Lisa M. Monteggia¹

Clinical studies consistently demonstrate that a single sub-psychomimetic dose of ketamine, an ionotropic glutamatergic NMDAR (*N*-methyl-*D*-aspartate receptor) antagonist, produces fast-acting antidepressant responses in patients suffering from major depressive disorder, although the underlying mechanism is unclear^{1–3}. Depressed patients report the alleviation of major depressive disorder symptoms within two hours of a single, low-dose intravenous infusion of ketamine, with effects lasting up to two weeks^{1–3}, unlike traditional antidepressants (serotonin re-uptake inhibitors), which take weeks to reach efficacy. This delay is a major drawback to current therapies for major depressive disorder and faster-acting antidepressants are needed, particularly for suicide-risk patients³. The ability of ketamine to produce rapidly acting, long-lasting antidepressant responses in depressed patients provides a unique opportunity to investigate underlying cellular mechanisms. Here we show that ketamine and other NMDAR antagonists produce fast-acting behavioural antidepressant-like effects in mouse models, and that these effects depend on the rapid synthesis of brain-derived neurotrophic factor. We find that the ketamine-mediated blockade of NMDAR at rest deactivates eukaryotic elongation factor 2 (eEF2) kinase (also called CaMKIII), resulting in reduced eEF2 phosphorylation and de-suppression of translation of brain-derived neurotrophic factor. Furthermore, we find that inhibitors of eEF2 kinase induce fast-acting behavioural antidepressant-like effects. Our findings indicate that the regulation of protein synthesis by spontaneous neurotransmission may serve as a viable therapeutic target for the development of fast-acting antidepressants.

We examined the acute effect of ketamine in wild-type C57BL/6 mice and detected notable behavioural responses in antidepressant-predictive tasks, including the forced swim test (FST), novelty-suppressed feeding (NSF) and learned helplessness (Supplementary Figs 1a–e and 2a–c)⁴. Ketamine also produced such responses in a sucrose consumption test, as well as in NSF and FST, after chronic mild stress, an animal model of depression (Supplementary Fig. 1f–i). To elucidate the mechanisms underlying the fast-acting antidepressant action of ketamine, we focused on FST, a test that is predictive of non-monoaminergic antidepressant efficacy⁴. We examined the time course of behavioural antidepressant effects in wild-type mice after a single, low-dose treatment with ketamine, (5*S*,10*R*)-(+)-5-methyl-10,11-dihydro-5*H*-dibenzo(a,d)cyclohepten-5,10-imine maleate (MK-801) or 3-((*R*)-2-carboxypiperazin-4-yl)-prop-2-enyl-1-phosphonic acid (CPP) (Fig. 1a–c). After either 30 min or 3 h, each NMDAR antagonist markedly reduced the immobility of mice in FST, when compared to vehicle-treated animals, indicating that NMDAR blockade produces fast-acting antidepressant responses. Notably, in our system, acute treatment with conventional antidepressants did not produce antidepressant-like FST responses (Supplementary Fig. 3), which may require multiple doses⁵. The effects of ketamine and CPP, but not of MK-801, persisted for 24 h (ref. 4) and ketamine's behavioural effect lasted for 1 week. Acute NMDAR-antagonist

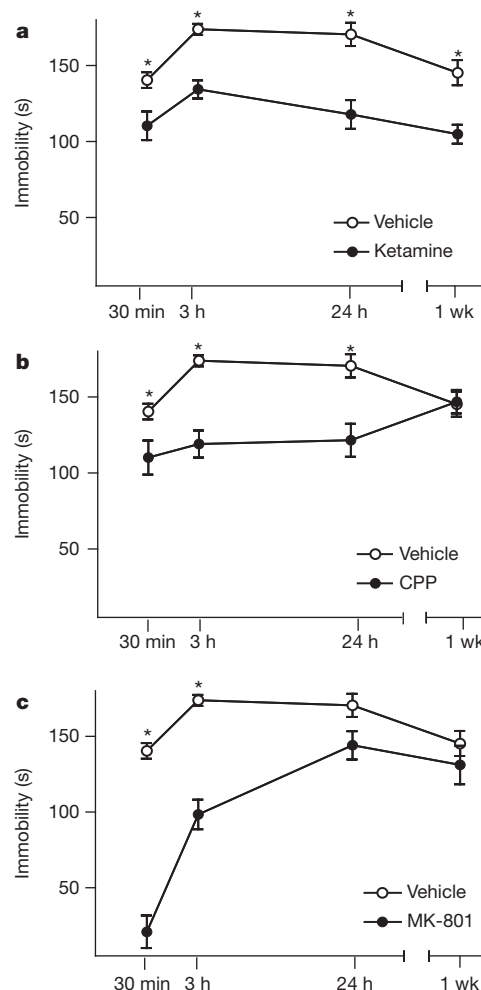


Figure 1 | Time course of NMDAR antagonist-mediated antidepressant-like behavioural effects. Mean immobility \pm s.e.m. of C57BL/6 mice in FST after acute treatment with ketamine, CPP or MK-801. Independent groups of mice were used at each time point and for each drug treatment, to avoid behavioural habituation. Analysis of variance (ANOVA) $F_{(3,27)} = 30.31$, $P < 0.0001$ for treatment groups; $F_{(3,27)} = 19.06$, $P < 0.0001$ for duration of response; $F_{(9,81)} = 9.32$, $P < 0.0001$ for treatment-duration interaction. Therefore, we examined treatment effects by time point. **a**, Ketamine (3.0 mg kg^{-1}) significantly reduced immobility, indicating an antidepressant-like response, at 30 min, 3 h, 24 h and 1 week, compared to vehicle treatment. **b**, CPP (0.5 mg kg^{-1}) significantly reduced immobility at 30 min, 3 h and 24 h, compared to vehicle treatment. **c**, MK-801 (0.1 mg kg^{-1}) produced significant decreases in immobility at 30 min and 3 h compared to vehicle treatment. $n = 10$ mice per group per time point; *, $P < 0.05$. Here and in all figures, error bars represent s.e.m.

¹Department of Psychiatry, University of Texas Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, Texas 75390-9111, USA. ²Department of Neuroscience, University of Texas Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, Texas 75390-9111, USA.

treatment produced no alterations in hippocampal-dependent learning (Supplementary Fig. 1d) or locomotor activity (Supplementary Fig. 4). These drugs have short half-lives (about 2–3 h)^{6–8}, indicating that sustained NMDAR-antagonist-induced antidepressant responses are due to synaptic plasticity, not to persistent blockade of receptors.

Brain-derived neurotrophic factor (BDNF) is linked to traditional antidepressant action; BDNF expression in the hippocampus is increased by antidepressants⁹ and BDNF deletion in the hippocampus attenuates antidepressant behavioural responses^{10–12}. Moreover, intra-ventricular or intrahippocampal BDNF infusion causes rapid, sustained antidepressant-like effects, lasting 3–6 days in FST^{13,14}. To examine whether the antidepressant-like response to ketamine is mediated through BDNF, we administered ketamine to inducible *Bdnf*-knockout mice¹⁰, then observed FST behaviour. After 30 min, ketamine-treated wild-type littermate controls showed significant reductions in immobility, indicating antidepressant-like responses, when compared to vehicle-treated controls (Fig. 2a). However, ketamine did not produce antidepressant-like effects in *Bdnf*-knockouts, indicating that fast-acting antidepressant responses require BDNF. After 24 h, ketamine significantly reduced immobility in controls, but not in *Bdnf* knockouts (Fig. 2a), indicating that ketamine's sustained effects depend on BDNF. To validate this link between NMDAR antagonists and BDNF-mediated antidepressant responses, MK-801 was administered to *Bdnf* knockouts or controls. After 30 min, MK-801 significantly reduced FST immobility in controls, but had no effect in *Bdnf* knockouts (Supplementary Fig. 6). MK-801 did not affect FST behaviour after 24 h (Supplementary Fig. 6), as previously demonstrated (Fig. 1c). We next generated postnatal conditional¹⁵ knockouts in neurotrophic tyrosine kinase receptor, type 2 (*Ntrk2*, also called *TrkB*) and found that these mice were insensitive to ketamine's antidepressant-like effects in FST and NSF (Supplementary Fig. 5a, b). To confirm TrkB engagement, we examined receptor autophosphorylation and found increased TrkB activation after NMDAR antagonist treatment (Supplementary Fig. 5c).

To determine whether NMDAR antagonists alter *Bdnf* expression in the hippocampus, wild-type mice were treated acutely with vehicle, ketamine or MK-801. Quantitative RT-PCR analysis of the coding exon of *Bdnf* showed that *Bdnf* mRNA expression was unaltered by ketamine or MK-801 at either 30 min or 24 h after treatment (Supplementary Fig. 7a). Contrastingly, western blot and ELISA analyses showed that BDNF protein levels were markedly increased at 30 min, but not at 24 h, after NMDAR antagonist treatment (Fig. 2b and Supplementary Fig. 7b). Moreover, the acute effects of ketamine on BDNF extended to its precursor, proBDNF (Supplementary Fig. 7c). These data indicate that rapid increases in BDNF protein translation, not transcription, are necessary for fast-onset antidepressant responses. However, continued BDNF protein upregulation does not underlie ketamine's long-term behavioural effects.

To study further the roles of translation and transcription in ketamine's antidepressant-like effects, we examined FST behaviour in mice treated with the protein synthesis inhibitor anisomycin¹⁶ or with the RNA polymerase inhibitor actinomycin D¹⁷, which block their respective processes by about 80% within 2 h. We pretreated mice with anisomycin or actinomycin D before treating them with ketamine (Fig. 2c). Anisomycin prevented the ketamine-induced rapid behavioural responses seen at 30 min in FST and NSF paradigms, indicating a dependence on new protein synthesis (Fig. 2d and Supplementary Fig. 8a, b). Anisomycin also prevented ketamine's long-term effect on FST (24 h), indicating that rapid protein translation was involved in sustained antidepressant-like responses (Fig. 2e). We found that the synthesis of both mature BDNF and proBDNF in the hippocampus was sensitive to anisomycin treatment (Supplementary Fig. 8c, d). However, actinomycin D did not affect ketamine's antidepressant-like effect on FST at either time point, indicating that it is independent of new gene expression (Supplementary Fig. 9b, c). To confirm that actinomycin D crossed the blood–brain barrier, we examined *Bdnf* mRNA expression

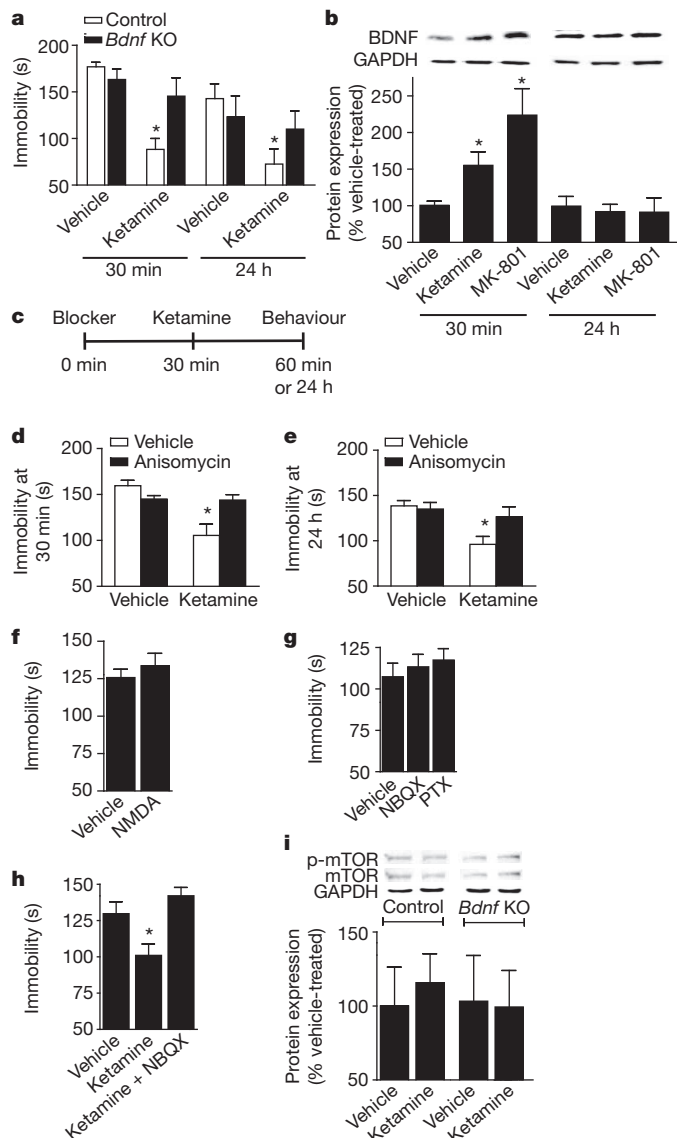


Figure 2 | BDNF translation in the antidepressant effects of NMDAR antagonists. **a**, Immobility in FST after acute treatment with ketamine (3.0 mg kg^{-1}). At 30 min, ANOVA $F_{(1,35)} = 17.13$, $P = 0.0002$ for drug; $F_{(1,35)} = 7.57$, $P = 0.0093$ for genotype–drug interaction; multiple comparisons with *t*-test, $*$, $P < 0.05$. At 24 h, in a separate cohort, ANOVA $F_{(1,29)} = 3.77$, $P = 0.0619$ for treatment; multiple comparisons with *t*-test, $*$, $P < 0.05$. $n = 7$ –12 mice per group. **b**, Densitometric analysis of BDNF (normalized to GAPDH) in the hippocampus after treatment with vehicle (control), ketamine (3.0 mg kg^{-1}) or MK-801 (0.1 mg kg^{-1}). At 30 min, ANOVA $F_{(2,12)} = 6.77$, $P = 0.0108$ for treatment, Bonferroni post hoc test, $*$, $P < 0.05$. At 24 h, no significant differences were seen ($n = 5$ –6 per group). **c**, Protocol for experiments using the blockers anisomycin and actinomycin D. **d**, Immobility at 30 min after anisomycin treatment. ANOVA $F_{(1,34)} = 11.83$, $P = 0.0016$ for treatment and $F_{(1,34)} = 10.91$, $P = 0.0023$ for treatment–inhibitor interaction; multiple comparisons, $*$, $P < 0.05$ ($n = 8$ –10 per group). **e**, Immobility at 24 h after anisomycin treatment. ANOVA $F_{(1,31)} = 9.34$, $P = 0.0046$ for treatment; multiple comparisons, $*$, $P < 0.05$ ($n = 8$ –10 per group). **f**, Immobility of wild-type mice given vehicle or NMDA (75 mg kg^{-1}), tested 30 min later in FST. **g**, Immobility of mice given NBQX (10 mg kg^{-1}) or picrotoxin (1 mg kg^{-1}), tested 30 min later in FST. **h**, Immobility of mice given vehicle, ketamine (3.0 mg kg^{-1}) or ketamine + NBQX (10 mg kg^{-1}) and tested 30 min later in FST. ANOVA $F_{(2,26)} = 8.226$, $P < 0.0019$; Bonferroni post hoc analysis shows that the ketamine effect is reversed by NBQX, $*$, $P < 0.05$. **i**, Densitometric analysis of phosphorylated mTOR (normalized to mTOR) in the hippocampus 30 min after treatment with vehicle or ketamine.

in drug-treated animals and found decreased *Bdnf* transcription in the hippocampus (Supplementary Fig. 9a). Taken together, these findings indicate that rapid, transient translation of BDNF is required for ketamine's fast-acting and long-lasting antidepressant-like behavioural effects and that long-term antidepressant responses may be due to alterations in synaptic plasticity, initiated by transient increases in BDNF translation.

We observed increased levels of BDNF protein in the cortex, but not in the nucleus accumbens, 30 min after acute administration of ketamine or MK-801 (Supplementary Fig. 10a, b). We further investigated whether NMDAR antagonism affected proteins other than BDNF. We found an increased level of activity-regulated cytoskeletal-associated protein (ARC) in the hippocampus (sensitive to anisomycin treatment; Supplementary Fig. 8e) but there was no increase in HOMER or GRIA1 (also known as GLUR1), nor in the phosphorylation of ribosomal protein S6 kinase (Supplementary Fig. 10c–f). Additionally, these proteins remained unaltered in the cortex after acute treatment with NMDAR antagonists (Supplementary Fig. 11a–e).

Synaptic plasticity and ensuing learning processes are often mediated by NMDAR-activation-driven protein translation, but antidepressant-like effects require protein translation induced by NMDAR blockade. To resolve this paradox, we turned to recent evidence that NMDAR blockade by MK-801 or 2-amino-5-phosphonopentanoic acid (AP5) without neuronal activity, augments protein synthesis through eEF2 dephosphorylation (activation). eEF2 is a critical catalytic factor for ribosomal translocation during protein synthesis¹⁸. In this model, resting NMDAR activity causes sustained activation of eEF2 kinase (eEF2K, or CamKIII), which phosphorylates eEF2, effectively halting translation, whereas acute NMDAR blockade at rest (in the absence of action potentials) attenuates eEF2 phosphorylation, allowing the translation of target transcripts.

To evaluate this model, we tested whether excess synaptic glutamate, possibly elicited by NMDAR blockade, was responsible for the behavioural effects of ketamine. Acute NMDA administration did not alter FST behaviour (Fig. 2f), as previously demonstrated¹⁹, but it increased ARC expression (Supplementary Fig. 10i), indicating that excess glutamate does not elicit rapid behavioural antidepressant effects. To define the role of neuronal activity in antidepressant behavioural effects, we tested whether NBQX, a blocker of α -amino-3-hydroxy-5-methyl-4-isoxazole propionic acid (AMPA) channels that reduces neuronal activity, or picrotoxin, a blocker of GABA (γ -aminobutyric acid) channels that increases activity, affected FST behaviour^{4,20}. Acute systemic treatment with these drugs did not affect FST behaviour (Fig. 2g) or BDNF synthesis, though picrotoxin enhanced ARC expression in the hippocampus (Supplementary Fig. 10g, h). However, when co-applied with ketamine, NBQX abolished behavioural antidepressant-like responses in FST (Fig. 2h), as previously described⁴. These data indicate that behavioural antidepressant effects are not elicited by alterations in evoked neurotransmission, but require ketamine-mediated augmentation of AMPA-receptor activation.

Recent evidence indicates that cortical mTOR signalling underlies ketamine-mediated antidepressant responses²¹. We investigated whether the rapid behavioural antidepressant effects of ketamine required mTOR activation, and whether this signalling was downstream of BDNF. Regulation of phosphorylated mTOR was not detected after acute administration of ketamine in control or *Bdnf*-knockout hippocampal tissue (Fig. 2i), nor in wild-type cortical tissue (Supplementary Fig. 11d). In earlier work, rapamycin prevented ketamine-mediated antidepressant responses; however, the link between rapamycin and antidepressant-like effects is equivocal²². We tested whether pre-treatment with rapamycin could block acute ketamine-mediated FST behaviour. Thirty minutes after ketamine administration, wild-type mice showed antidepressant responses unaffected by rapamycin treatment (Supplementary Fig. 11h). Rapamycin reduced the phosphorylation of ribosomal protein S6 kinase in the cortex and hippocampus (Supplementary Fig. 11f, g), indicating that the rapamycin had penetrated

brain tissue. The earlier study examined molecular effects 2 h after drug treatment, or behavioural effects 24 h after drug treatment²¹; therefore mTOR's role in the antidepressant effect of ketamine may be one of maintenance rather than rapid induction.

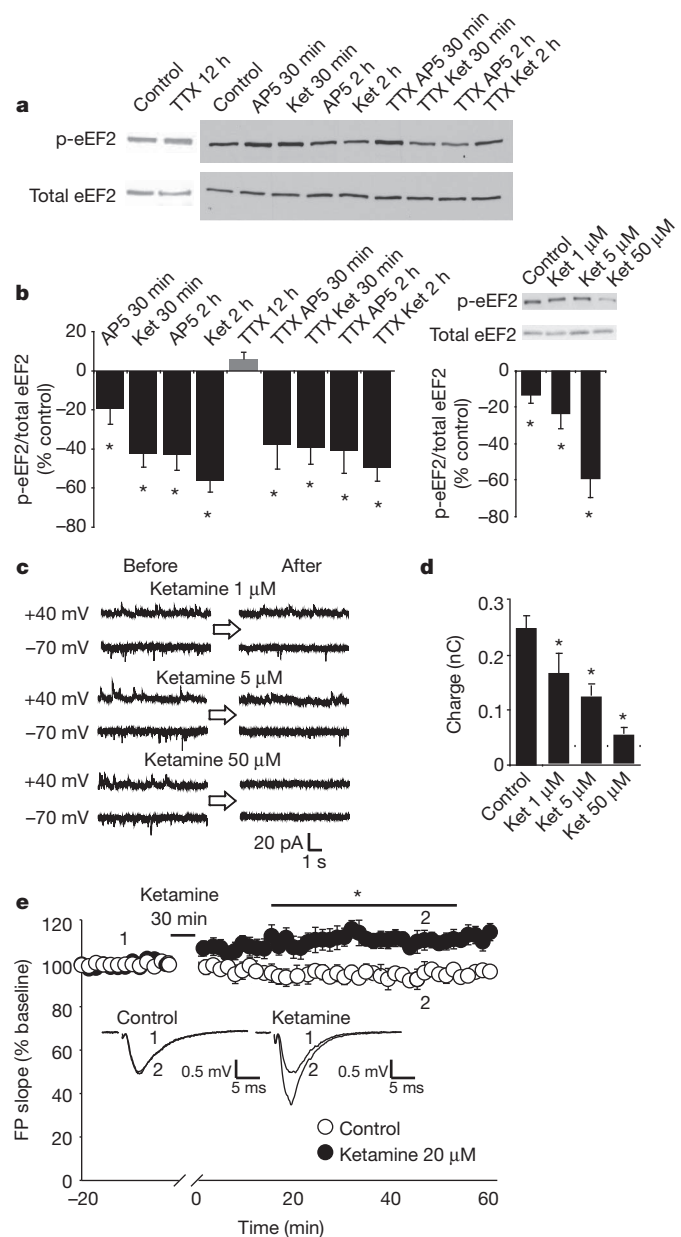
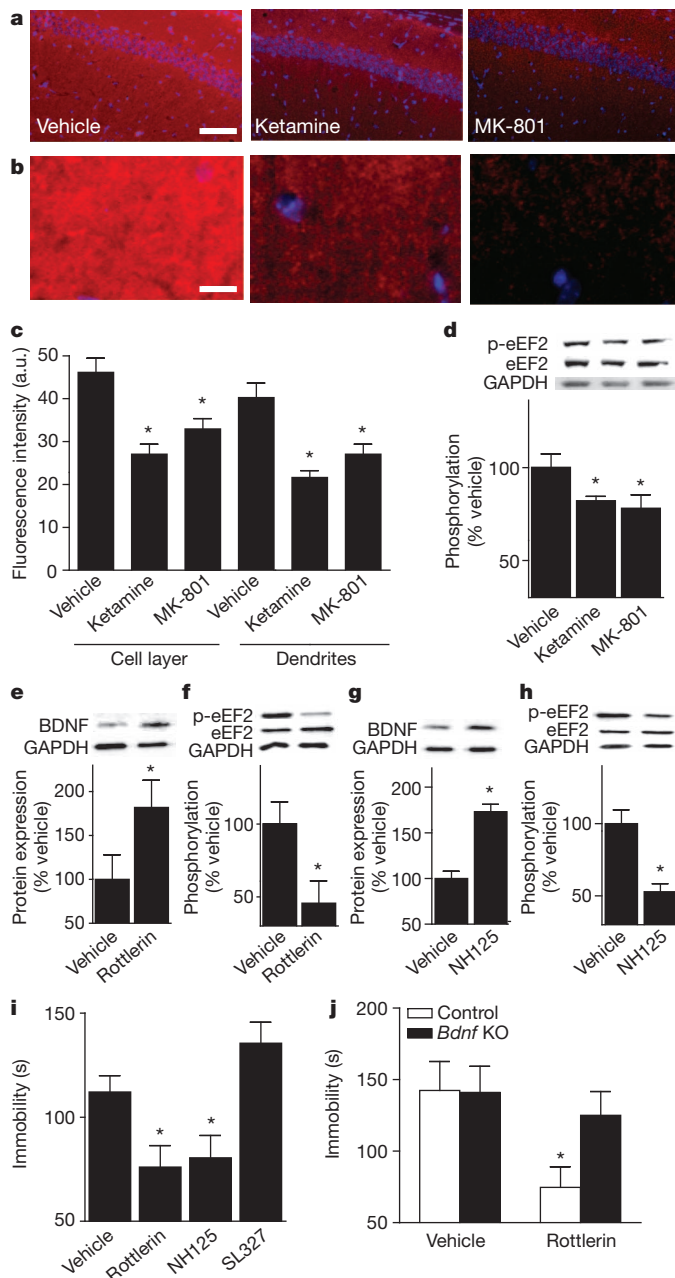


Figure 3 | Ketamine blocks NMDAR spontaneous activity, reduces the level of eEF2 phosphorylation and strengthens synaptic responses.

a, Representative Western blots showing eEF2 phosphorylation (p-eEF2) in hippocampal primary cultures. Ket, ketamine; TTX, tetrodotoxin. **b**, Densitometric analysis of p-eEF2 normalized to total eEF2 (left panel). Data are expressed as mean percentage \pm s.e.m. Tetrodotoxin alone does not alter p-eEF2, whereas AP5 or ketamine, with or without tetrodotoxin, significantly reduce the level of p-eEF2, as assessed by *t*-test (*, $P < 0.05$). Right panel: application of 1 μ M, 5 μ M or 50 μ M ketamine causes dose-dependent decreases in p-eEF2, as assessed by *t*-test (*, $P < 0.05$). **c**, Representative traces of NMDAR spontaneous activity after application of 1 μ M, 5 μ M or 50 μ M ketamine. **d**, Quantification of charge transfer (10 s) reveals significant effects, as assessed by *t*-test, for all ketamine concentrations compared to controls ($n = 6-16$; *, $P < 0.05$). **e**, Field-potential (FP) slopes are plotted as a function of time. Representative field-potential traces (average 2 min) are shown during baseline (1) and at 45 min (2). The asterisk refers to significantly different field-potential values (*, $P < 0.05$). For statistical analysis, we used two-way repeated ANOVA with Bonferroni post hoc analysis. The drug-time interaction was significant ($F_{(143,1430)} = 6.723$, $P < 0.001$).

To determine whether ketamine inhibits NMDA-receptor-mediated spontaneous miniature excitatory postsynaptic currents (NMDAR-mEPSCs)^{3,24} when at rest, and whether it regulates eEF2 phosphorylation, we tested its impact on hippocampal neurons *in vitro*. After ketamine perfusion (1 μ M, 5 μ M or 50 μ M), we recorded NMDAR-mEPSCs (Fig. 3c) and detected a significant decrease within minutes, similar to the effect of AP5 (ref. 23). Moreover, protein extracts from ketamine-treated neurons showed decreased eEF2 phosphorylation compared to vehicle-treated cultures, indicating that ketamine, in the absence of neuronal activity, dose-dependently leads to eEF2 de-phosphorylation, permitting protein synthesis (Fig. 3a, b). Additionally, we evaluated ketamine's effects on hippocampal field potentials. Acute application of ketamine (20 μ M, at rest) potentiated the synaptic responses subsequently evoked in hippocampal slices (Fig. 3e), further showing that increased AMPA-mediated neurotransmission underlies ketamine's antidepressant-like behavioural effects. This result is consistent with findings regarding BDNF-dependent and protein-synthesis-dependent synaptic plasticity²⁵.



To examine whether the fast-acting antidepressant response is mediated via eEF2, we administered ketamine or MK-801 to wild-type mice and analysed eEF2 phosphorylation. Within 30 min, ketamine and MK-801 led to rapid decreases in the level of phosphorylated eEF2 in the hippocampus (Fig. 4a–c and Supplementary Figs 12 and 13), detected by immunostaining and western blot analysis (Fig. 4d). However, cortical levels of phosphorylated eEF2 were unaltered after acute NMDAR-antagonist treatment (Supplementary Fig. 11f). To examine whether eEF2K inhibition alters BDNF protein expression *in vivo*, the eEF2K inhibitors rottlerin or 1-hexadecyl-2-methyl-3-(phenylmethyl)-1H-imidazolium iodide (NH125) were administered to wild-type mice and the mice were killed 30 min later. Rottlerin and NH125 produced significantly increased BDNF protein expression (Fig. 4e, g), with corresponding significant decreases in phosphorylated eEF2 in the hippocampus (Fig. 4f, h). To assess directly whether eEF2K inhibition is sufficient to mediate fast-acting antidepressant-like responses, wild-type mice were treated with rottlerin or NH125 and then examined in FST. Both rottlerin and NH125 produced significant decreases in FST immobility at 30 min (Fig. 4i), a timescale similar to that of the effects of NMDAR antagonists, indicating that fast-acting behavioural effects are mediated through eEF2K inhibition. To test whether mitogen-activated protein kinase (MAPK), a regulator of protein translation during neural activity, affects FST behaviour, we treated wild-type mice with the inhibitor SL327. This treatment reduced MAPK phosphorylation in hippocampal tissue (Supplementary Fig. 10j), but did not affect FST behaviour (Fig. 4i), indicating that antidepressant-like effects are specific to eEF2K inhibition during resting spontaneous glutamatergic signalling. We found that an acute dose of rottlerin or NH125 did not affect locomotor activity, but that antidepressant-related behavioural effects were long-lasting (Supplementary Fig. 14a–f). To validate the finding that antidepressant effects after eEF2K inhibition were mediated through BDNF, we administered rottlerin to *Bdnf*-knockout mice and tested FST behaviour. Like NMDAR antagonists, rottlerin was ineffective in *Bdnf* knockouts, showing that increased *Bdnf* expression upon eEF2K inhibition is required to produce antidepressant-like behavioural responses (Fig. 4j).

Our data support the hypothesis that ketamine produces rapidly acting antidepressant-like behavioural effects through inhibition of spontaneous NMDAR-mEPSCs, leading to decreased eEF2K activity, thus permitting rapid increases in BDNF translation (Supplementary Fig. 15) which may, in turn, exert strong influences on presynaptic or

Figure 4 | Rapid antidepressant-like behaviour is mediated by decreased p-eEF2 and increased BDNF translation. **a**, Images of CA1 pyramidal and stratum radiatum layers after acute treatment with vehicle, ketamine or MK-801. Scale bar, 100 μ m; red, p-eEF2; blue, DAPI. **b**, Magnification of stratum radiatum; scale bar, 20 μ m. **c**, ImageJ analysis of average fluorescence intensity (a.u., arbitrary units). ANOVA on cell layer, $F_{(2,23)} = 13.13$, $P = 0.0002$ for treatment; ANOVA on dendrites, $F_{(2,23)} = 14.06$, $P = 0.0001$ for treatment ($n = 4$ per group; *, $P < 0.05$). **d**, Densitometric analysis of p-eEF2 normalized to total eEF2 in the hippocampus after treatment with NMDAR antagonists. ANOVA $F_{(2,23)} = 3.183$, $P = 0.03$ for treatment ($n = 8$ per group). **e–h**, Densitometric analyses of BDNF and p-eEF2. Significant increases are seen in hippocampal BDNF protein levels (normalized to GAPDH) with rottlerin (5 mg kg⁻¹) versus vehicle (**e**), and with NH125 (5 mg kg⁻¹) versus vehicle (**g**) (t -tests, *, $P < 0.05$). Significant decreases are seen in p-eEF2 (normalized to total eEF2) with rottlerin versus vehicle (**f**) and NH125 versus vehicle (**h**) (t -tests, *, $P < 0.05$). **i**, Immobility in FST of wild-type mice given acute rottlerin (5 mg kg⁻¹) or NH125 (5 mg kg⁻¹). ANOVA $F_{(3,44)} = 8.13$, $P = 0.0002$ for treatment; Bonferroni post hoc analysis shows significance with rottlerin or NH125 versus vehicle (*, $P < 0.05$), but not with the MAPK inhibitor SL327 (10 mg kg⁻¹). **j**, Immobility of *Bdnf*-knockout mice or littermate controls given acute rottlerin (5 mg kg⁻¹) and tested 30 min later in FST. ANOVA $F_{(1,19)} = 5.77$, $P = 0.0267$ for treatment; Bonferroni post hoc analysis for rottlerin versus vehicle-treated controls (*, $P < 0.05$; $n = 5–7$ per group).

postsynaptic efficacy^{26,27}. We found that fast-acting antidepressant-like effects cannot be elicited by disinhibition of behavioural circuitry, or by evoked neurotransmission, but must rely on enhanced neurotransmission after NMDAR-antagonist-induced plasticity, occurring at rest¹⁸. The observation of behavioural effects mediated through spontaneous neurotransmission provides the first evidence that tonic resting neurotransmission is involved in behaviour, and supports the notion that spontaneous and evoked forms of glutamatergic signalling are segregated^{18,23,28,29}. These data demonstrate that eEF2K inhibition, resulting in de-suppression of protein translation, is sufficient to produce antidepressant-like effects, implicating eEF2K inhibitors as potential novel major depressive disorder treatments with rapid onset. Moreover, our results show that synaptic translational machinery may serve as a viable therapeutic target for the development of faster-acting antidepressants.

METHODS SUMMARY

Behavioural studies were performed using adult male C57BL/6 wild-type or mutant mouse strains, maintained as previously described^{10,15}. All drugs were administered via intraperitoneal injection. Antidepressant-like behaviour was assessed using the forced swim test, as previously described⁴. Briefly, animals were placed in a cylinder of water at 22–24 °C for 6 min and immobility was measured during the last 4 min of the test. Molecular studies consisted of western blot analysis or quantitative PCR performed on whole-cell lysates from medial prefrontal cortex or anterior hippocampus. Electrophysiological studies were performed as previously described in cultured neurons (whole-cell recordings²³) or in hippocampal slices (field recordings¹⁰).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 25 October 2010; accepted 21 April 2011.

Published online 15 June 2011.

- Zarate, C. A. Jr *et al.* A randomized trial of an *N*-methyl-D-aspartate antagonist in treatment-resistant major depression. *Arch. Gen. Psychiatry* **63**, 856–864 (2006).
- Berman, R. M. *et al.* Antidepressant effects of ketamine in depressed patients. *Biol. Psychiatry* **47**, 351–354 (2000).
- Price, R. B., Nock, M. K., Charney, D. S. & Mathew, S. J. Effects of intravenous ketamine on explicit and implicit measures of suicidality in treatment-resistant depression. *Biol. Psychiatry* **66**, 522–526 (2009).
- Maeng, S. *et al.* Cellular mechanisms underlying the antidepressant effects of ketamine: role of α -amino-3-hydroxy-5-methylisoxazole-4-propionic acid receptors. *Biol. Psychiatry* **63**, 349–352 (2008).
- Detke, M. J., Johnson, J. & Lucki, I. Acute and chronic antidepressant drug treatment in the rat forced swimming test model of depression. *Exp. Clin. Psychopharmacol.* **5**, 107–112 (1997).
- Sinner, B. & Graf, B. M. Ketamine. *Handb. Exp. Pharmacol.* **182**, 313–333 (2008).
- Schwartz, P. H. & Wasterlain, C. G. Cardiac arrest and resuscitation alters the pharmacokinetics of MK-801 in the rat. *Res. Commun. Chem. Pathol. Pharmacol.* **73**, 181–195 (1991).
- Kristensen, J. D., Hartvig, P., Karlsten, R., Gordh, T. & Halldin, M. CSF and plasma pharmacokinetics of the NMDA receptor antagonist CPP after intrathecal, extradural and i.v. administration in anaesthetized pigs. *Br. J. Anaesth.* **74**, 193–200 (1995).
- Chen, B., Dowlatsahi, D., MacQueen, G. M., Wang, J. F. & Young, L. T. Increased hippocampal BDNF immunoreactivity in subjects treated with antidepressant medication. *Biol. Psychiatry* **50**, 260–265 (2001).
- Monteggia, L. M. *et al.* Essential role of brain-derived neurotrophic factor in adult hippocampal function. *Proc. Natl Acad. Sci. USA* **101**, 10827–10832 (2004).
- Berton, O. *et al.* Essential role of BDNF in the mesolimbic dopamine pathway in social defeat stress. *Science* **311**, 864–868 (2006).
- Adachi, M., Barrot, M., Autry, A. E., Theobald, D. & Monteggia, L. M. Selective loss of brain-derived neurotrophic factor in the dentate gyrus attenuates antidepressant efficacy. *Biol. Psychiatry* **63**, 642–649 (2008).
- Shirayama, Y., Chen, A. C., Nakagawa, S., Russell, D. S. & Duman, R. S. Brain-derived neurotrophic factor produces antidepressant effects in behavioral models of depression. *J. Neurosci.* **22**, 3251–3261 (2002).
- Hoshaw, B. A., Malberg, J. E. & Lucki, I. Central administration of IGF-I and BDNF leads to long-lasting antidepressant-like effects. *Brain Res.* **1037**, 204–208 (2005).
- Akbarian, S. *et al.* Brain-derived neurotrophic factor is essential for opiate-induced plasticity of noradrenergic neurons. *J. Neurosci.* **22**, 4153–4162 (2002).
- Lattal, K. M. & Abel, T. Different requirements for protein synthesis in acquisition and extinction of spatial preferences and context-evoked fear. *J. Neurosci.* **21**, 5773–5780 (2001).
- Capasso, A., Di Giannuario, A., Loizzo, A., Pieretti, S. & Sorrentino, L. Actinomycin D blocks the reducing effect of dexamethasone on amphetamine and cocaine hypermotility in mice. *Gen. Pharmacol.* **27**, 707–712 (1996).
- Sutton, M. A., Taylor, A. M., Ito, H. T., Pham, A. & Schuman, E. M. Postsynaptic decoding of neural activity: eEF2 as a biochemical sensor coupling miniature synaptic transmission to local protein synthesis. *Neuron* **55**, 648–661 (2007).
- Poleszak, E. *et al.* NMDA/glutamate mechanism of antidepressant-like action of magnesium in forced swim test in mice. *Pharmacol. Biochem. Behav.* **88**, 158–164 (2007).
- Fernandez, F. *et al.* Pharmacotherapy for cognitive impairment in a mouse model of Down syndrome. *Nature Neurosci.* **10**, 411–413 (2007).
- Li, N. *et al.* mTOR-dependent synapse formation underlies the rapid antidepressant effects of NMDA antagonists. *Science* **329**, 959–964 (2010).
- Cleary, C. *et al.* Antidepressive-like effects of rapamycin in animal models: Implications for mTOR inhibition as a new target for treatment of affective disorders. *Brain Res. Bull.* **76**, 469–473 (2008).
- Atasoy, D. *et al.* Spontaneous and evoked glutamate release activates two populations of NMDA receptors with limited overlap. *J. Neurosci.* **28**, 10151–10166 (2008).
- Espinosa, F. & Kavalali, E. T. NMDA receptor activation by spontaneous glutamatergic neurotransmission. *J. Neurophysiol.* **101**, 2290–2296 (2009).
- Tanaka, J. *et al.* Protein synthesis and neurotrophin-dependent structural plasticity of single dendritic spines. *Science* **319**, 1683–1687 (2008).
- Jakawich, S. K. *et al.* Local presynaptic activity gates homeostatic changes in presynaptic function driven by dendritic BDNF synthesis. *Neuron* **68**, 1143–1158 (2010).
- Lindskog, M. *et al.* Postsynaptic GluA1 enables acute retrograde enhancement of presynaptic function to coordinate adaptation to synaptic inactivity. *Proc. Natl Acad. Sci. USA* **107**, 21806–21811 (2010).
- Sutton, M. A. & Schuman, E. M. Partitioning the synaptic landscape: distinct microdomains for spontaneous and spike-triggered neurotransmission. *Sci. Signal.* **2**, pe19 (2009).
- Kavalali, E. T. *et al.* Spontaneous neurotransmission: an independent pathway for neuronal signaling? *Physiology (Bethesda)* **26**, 45–53 (2011).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. A. Mahgoub for assistance with the animal experiments, S. Birnbaum and A. Pettersen for assistance with the behavioural testing and members of the Monteggia and Kavalali laboratories for discussions and comments on the manuscript. This work was supported by grant MH070727 (L.M.M.), grant MH066198 (E.T.K.) and the Division of Basic Sciences Training Program at UT Southwestern Medical Center, T32 MH 76690-02 (A.E.A.). E.T.K. is an Established Investigator of the American Heart Association.

Author Contributions A.E.A. performed the behavioural experiments. A.E.A., M.A. and M.F.L. contributed to the molecular experiments. E.N. performed the electrophysiology experiments, E.S.N. performed the TrkB behavioural experiments and A.E.A. and P.-f.C. performed the statistical analyses. A.E.A. also made the figures and wrote the corresponding sections of the paper. E.T.K. and L.M.M. designed the study, supervised the experiments and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to L.M.M. (lisa.monteggia@utsouthwestern.edu) or E.T.K. (ege.kavalali@utsouthwestern.edu).

METHODS

Mouse. C57BL/6 male mice aged 6–8 weeks old were habituated to animal facilities for 1 week before behavioural testing. Mice were kept on a 12 h/12 h light/dark cycle and were given access to food and water *ad libitum*. Inducible *Bdnf* knockouts were generated from a trigenic cross of NSE-tTA, TetOp-Cre and floxed *Bdnf* mice, as previously described¹⁰. Conditional *Ntrk2*-knockout mice were made by crossing CamK-cre(93) (ref. 15) to floxed *Ntrk2* mice. For all behavioural testing, male mice were 2–4 months old and weight-matched, and groups were balanced by genotype. All animal procedures conformed to the guide for the care and use of laboratory animals and were approved by the institutional animal care and use committee at UT Southwestern Medical Center.

Drugs. All drugs were injected intraperitoneally. Concentrations were as follows: ketamine (Fort Dodge Animal Health) 3.0 mg kg⁻¹, MK-801 (Sigma) 0.1 mg kg⁻¹ and CPP (Sigma) 0.5 mg kg⁻¹ in 0.9% saline; anisomycin (Sigma) 100 mg kg⁻¹ (dissolved in HCl/saline, final pH 7.4); actinomycin D (Sigma) 0.5 mg kg⁻¹ in 5% ethanol; roflumetinol and NH125 (Sigma) 5 mg kg⁻¹ in 20–100% DMSO; SL327 (Sigma) 10 mg kg⁻¹ in 100% DMSO³⁰; NMDA (Sigma) 75 mg kg⁻¹, NBQX (Sigma) 10 mg kg⁻¹ and picrotoxin (Sigma) 1 mg kg⁻¹ in 0.9% saline; rapamycin (Sigma) 1.0 mg kg⁻¹ dissolved in 50% DMSO.

Sucrose consumption test. Group-housed mice were habituated to a 1% solution of sucrose in tap-water for 48 h. The mice were then habituated to water-deprivation periods of 4 h, 14 h and 19 h, followed by a 1 h exposure to the sucrose solution for 3 days with intervening access to normal drinking water. To assess individual sucrose intake, the group-housed mice were water-deprived overnight and then housed temporarily in a new cage. Each test mouse was placed in its home cage for 1 h with access to the 1% sucrose solution. The bottle of sucrose solution was weighed before and after the test to determine sucrose intake. A water test was performed in a similar manner the following day. Data are expressed as a percentage of sucrose to total volume consumed in both sucrose and water trials.

Elevated plus maze. Mice were placed in the centre of a plus maze (each arm 33 cm × 5 cm) that was elevated 1 m above the floor with two open arms and two closed arms (25-cm-tall walls on the closed arms) at 40 lx. The exploratory activity was monitored for 5 min with a video tracking system and the duration, in seconds, spent in the closed and open arms was recorded by EthoVision software.

Novelty-suppressed feeding. Briefly, group-housed animals were food-deprived for 24 h and then placed in a temporary home cage for 30 min. For the test, individual mice were placed in a 42 × 42 cm open-field arena at 40 lx. A single pellet of the mouse's normal food chow was placed in the centre of the open-field arena. Each animal was placed in a corner of the arena and allowed to explore for up to 10 min. The trial ended when the mouse chewed a part of the chow. The amount of food consumed in the home cage was taken as the weight of chow consumed in 5 min, as a control measure for appetite.

Context and cued fear conditioning. Fear conditioning was performed as previously described⁵. Briefly, mice were placed in individual chambers for 2 min, followed by a loud tone (90 dB) for 30 s, immediately followed by a 0.5 mA foot-shock for 2 s. After 1 min, mice received a second pairing of tone and footshock, as described. Mice were placed in home cages until 24 h later, when the mice were placed back in the same boxes without a tone or shock. The amount of time that the animal spent freezing was scored by an observer blind to genotype. Freezing behaviour was defined as no movement except for respiration. Four hours later, mice were placed in a novel environment with no tone or shock for 3 min, followed by 3 min of the tone to assess cue-dependent fear conditioning. Again, time spent freezing was recorded as described¹⁰.

Learned helplessness. Mice were trained on one side of a two-chamber shuttlebox (MedAssociates) with the door closed for 1 h, receiving 120 variable-interval shocks (18–44 s, average 30 s; 0.35 mA for 2 s) on 2 training days. On the test day, the door was raised at the onset of the shock and the shock ended either when the mouse stepped through to the other side of the shuttlebox or after 25 s. Latency to step through the door and the number of escape failures were recorded for 15 trials.

Locomotor activity. Mice were placed in cages and locomotor activity was recorded for 1 h under red light by photocell beams linked to computer acquisition software (San Diego Instruments).

Forced swim test. The forced swim test (FST) was performed as previously described¹². This test is sensitive to conventional antidepressant treatment³¹ as well as to non-monoaminergic antidepressants⁴. Mice were placed for 6 min in a 4 l Pyrex glass beaker containing 3 l of water at 24 ± 1 °C. Water was changed between subjects. All test sessions were recorded by a video camera positioned on the side of the beaker. The videotapes were analysed and scored by an observer blind to group assignment during the last 4 min of the 6 min trial. A decrease in immobility time indicates an antidepressant-like response.

Chronic mild stress. Stressed mice were subjected to two randomly selected mild stressors per day, of variable duration (1–12 h), for 28 days. Stressors included

water deprivation, 45° cage-tilt, food deprivation, exposure to rat faeces, cage overcrowding, wet bedding, overnight illumination, dark exposure during normal light cycle, cold bedding, acoustic disturbance (120 dB), strobe lights and cage-mate rotation. Stressors were not applied within 8 h of behavioural testing.

Time course experiments. Separate cohorts of C57BL/6 adult male mice were injected intraperitoneally with vehicle or the NMDAR antagonists ketamine (3.0 mg kg⁻¹), MK-801 (0.1 mg kg⁻¹) or CPP (0.5 mg kg⁻¹) at 30 min, 3 h, 24 h or 1 week before FST (*n* = 10 per group). The drug doses were chosen on the basis of previous literature demonstrating an antidepressant-like response in mouse models⁴.

Anisomycin and actinomycin D experiments. Separate cohorts of C57BL/6 adult male mice were injected intraperitoneally with either vehicle or anisomycin (100 mg kg⁻¹), or with either saline or actinomycin D (0.5 mg kg⁻¹), 1 h before FST. Thirty minutes before testing, mice received either a saline or a ketamine injection (3.0 mg kg⁻¹) (*n* = 10 per group). For 24 h experiments, mice were given anisomycin (100 mg kg⁻¹) or saline 30 min before an injection of ketamine and were tested in the FST 1 day later.

Inducible *Bdnf*-knockout experiments. Separate cohorts of inducible *Bdnf* knockout adult male mice and wild-type littermate controls were subjected to FST either 30 min or 24 h after injection with saline, ketamine (3.0 mg kg⁻¹) or MK-801 (0.1 mg kg⁻¹) (*n* = 7–12 per group).

Quantitative RT-PCR. Fresh frozen anterior hippocampal slices (2 per mouse, ~1 mm thick) were dissected and total RNA was extracted using Trizol reagent (Invitrogen), according to manufacturer's instructions. Conditions for cDNA synthesis, amplification and primer sequences were described previously¹². The fold-change in *Bdnf* expression (coding exon) was normalized to GAPDH.

Protein quantification. Anterior hippocampal slices (2 per mouse, ~1 mm thick) were dissected from C57BL/6 mice that had received saline vehicle, ketamine (3.0 mg kg⁻¹) or MK-801 (0.1 mg kg⁻¹) injections, either 30 min or 24 h after injection. The slices were rapidly frozen and lysed in buffer containing protease inhibitors and phosphatase inhibitors. Total protein concentration was quantified by Bradford analysis. BDNF quantification was carried out by SDS-polyacrylamide gel electrophoresis. Primary antibodies for BDNF (Santa Cruz Biotechnology) and GAPDH (Cell Signaling) were used at dilutions of 1:200 and 1:10,000, and anti-rabbit secondary antibodies were used at 1:2,000 and 1:50,000, respectively. To measure phosphorylated eEF2 (p-eEF2, Thr 56) and total eEF2, primary antibodies were used at dilutions of 1:1,000 and anti-rabbit secondary antibodies were used at 1:2,000. Mouse anti-ARC (C7, Cell Signaling) was used at a primary dilution of 1:1,000 and secondary dilution of 1:2,000. Phospho-mTOR and total mTOR (Cell Signaling) were both used at primary dilutions of 1:500 and secondary dilutions of 1:10,000. GluR1 (Chemicon) was used at a primary dilution of 1:5,000 and secondary dilution, 1:2,000. Pan-HOMER antibody (Cell Signaling) was used at 1:5,000 with 1:2,000 dilutions for primary and secondary, respectively. Phospho-s6 kinase and total s6 kinase antibodies were used at 1:200 and 1:5,000 for primary dilutions, respectively, and both had secondary dilutions of 1:5,000 (Cell Signaling). Phospho-MAPK and total MAPK antibodies (Cell Signaling) were used at primary dilutions of 1:10,000 and 1:500 respectively and both had secondary dilutions of 1:2,000. Bands developed with enzymatic chemiluminescence (ECL) were exposed to film and films were analysed by ImageJ. BDNF was normalized to GAPDH bands, and p-eEF2 and total eEF2 bands were taken as a ratio of GAPDH-normalized values.

Immunohistochemistry. C57BL/6 mice were treated intraperitoneally with saline, ketamine (3.0 mg kg⁻¹) or MK-801 (0.1 mg kg⁻¹) and killed 30 min later. The protocol is adapted from a previous study³². Brains were fresh-dissected and fixed for 72 h in ice-cold 4% paraformaldehyde. Brains were cryoprotected for 2 or more hours in 20% glycerol, sectioned on a freezing microtome at 30 µm and preserved in 1× PBS with 0.01% sodium azide. Floating sections were washed in 2× SSC, followed by antigen-unmasking in 50:50 acetone:methanol, performed at 4 °C. Sections were rinsed and endogenous peroxidase activity was quenched in 1% H₂O₂ for 30 min. Sections were rinsed in 2× SSC with 0.05% Tween-20. Tissue was blocked for 30 min in 3% normal goat serum diluted in 2× SSC/0.05% Tween, followed by primary antibody, rabbit anti-p-eEF2 (diluted 1:100 in blocking solution; Cell Signaling Technology), and incubation for 48 h at 4 °C. After rinsing in 2× SSC, a horseradish-peroxidase-labelled secondary antibody at 1:200 was applied and the signal was amplified using the tyramide amplification signal system (Perkin Elmer). Slides were counterstained with 4',6'-diamidino-2-phenylindole (DAPI), mounted on superfrost plus slides, dried for 2 h and mounted in DPX mountant.

ELISA. A high-sensitivity enzyme-linked immunosorbent assay was used to assess BDNF levels, as per manufacturer's instructions (Promega). Briefly, hippocampal lysates were prepared in the recommended buffer, diluted 1:4 in 1× PBS and acid-treated as instructed by the manufacturer. A 96-well plate (Nunc) was coated overnight in carbonate coating buffer, blocked in the provided sample buffer for

2 h at 26 °C and treated with recombinant human BDNF antibody for 2 h at 26 °C. Acid-treated samples and provided standards were added to the plate in duplicate. Wells were then treated for 1 h at RT with anti-IgY conjugated to horseradish peroxidase and colour was developed with the provided 3,3',5,5'-tetramethylbenzidine (TMB) solution for 10 min, then stopped with 1 M HCl. Absorbance of wells was measured at 450 nm. BDNF concentration was determined by comparing the mean absorbance of the duplicate samples to the standards. BDNF concentration was then normalized to total protein content and expressed as pg of BDNF per μ g of total protein.

Cell culture. Dissociated hippocampal cultures were prepared as previously described³³. Briefly, whole hippocampi were dissected from Sprague–Dawley rats at postnatal day 0–3. Tissue was trypsinized (10 mg ml⁻¹ trypsin) for 10 min at 37 °C, mechanically dissociated by pipetting and plated on Matrigel-coated coverslips. Cytosine arabinoside (4 μ M, Sigma) was added at day 1 *in vitro* and the concentration of cytosine arabinoside was reduced to 2 μ M at day 4 *in vitro*. All experiments were performed on cultures at day 14–21.

Cell culture recordings. Whole-cell patch-clamp recordings were performed on hippocampal pyramidal neurons. Data were acquired using a MultiClamp 700B amplifier and Clampex 9.0 software (Molecular Devices). Recordings were filtered at 2 kHz and sampled at 200 μ s. A modified Tyrode's solution containing 150 mM NaCl, 4 mM KCl, 2 mM MgCl₂, 2 mM CaCl₂, 10 mM glucose and 10 mM HEPES, pH 7.4, was used as external bath solution. The pipette-internal solution contained 115 mM CsMeSO₃, 10 mM CsCl, 5 mM NaCl, 10 mM HEPES, 0.6 mM EGTA, 20 mM tetraethylammonium chloride, 4 mM Mg-ATP, 0.3 mM Na₃GTP, pH 7.35, and 10 mM QX-314 (*N*-(2,6-dimethylphenylcarbamoylmethyl)-triethylammonium bromide), 300 mosM. Series resistance was 10–30 m Ω . To record and isolate NMDAR-mEPSCs, the MgCl₂ concentration was reduced to 0.1 mM and 2,3-dihydroxy-6-nitro-7-sulfamoyl-benzo(f)quinoxaline-2,3-dione (NBQX; 10 μ M, Sigma) and picrotoxin (50 μ M; Sigma) were added to the bath solution to block AMPA-receptor-mediated excitatory currents and GABA (γ -aminobutyric acid) receptor-mediated inhibitory currents, respectively. The baseline for the analysis of NMDAR-mEPSCs was automatically determined as the average current level of silent episodes during a recording. The events were selected at a minimum threshold of 4 pA and the area under current deflection was calculated to quantify charge transfer¹⁸.

Field recordings. Field recordings were made from hippocampal slices from Sprague–Dawley rats obtained from Charles River Laboratories. Slices (400 μ m) were prepared from rats at 15–25 days old. Rats were anesthetized with euthasol

(50 mg kg⁻¹) and decapitated soon after the disappearance of corneal reflexes. The brain was removed, dissected and then sliced using a vibratome (1000 Plus) in ice-cold dissection buffer containing 2.6 mM KCl, 1.25 mM NaH₂PO₄, 26 mM NaHCO₃, 0.5 mM CaCl₂, 5 mM MgCl₂, 212 mM sucrose and 10 mM dextrose. Area CA3 was surgically removed from each slice immediately after sectioning. The slices were transferred into a reservoir chamber filled with artificial cerebrospinal fluid (ACSF) containing 124 mM NaCl, 5 mM KCl, 1.25 mM NaH₂PO₄, 26 mM NaHCO₃, 2 mM CaCl₂, 2 mM MgCl₂ and 10 mM dextrose. Slices were allowed to recover for 2–3 h at 30 °C. ACSF and dissection buffer were equilibrated with 95% O₂ and 5% CO₂. For recording, slices were transferred to a submerged recording chamber, maintained at 30 °C, and perfused continuously with ACSF at a rate of 2–3 ml min⁻¹. Field potentials were recorded with extracellular recording electrodes (1 M Ω) filled with ACSF and placed in the stratum radiatum of area CA1. Field potentials were evoked by monophasic stimulation (duration, 200 μ s) of Schaffer collateral/commissural afferents with a concentric bipolar tungsten-stimulating electrode (Frederick Haer Company). Stable baseline responses were collected every 30 s using a stimulation intensity of 10–30 μ A, yielding 50–60% of the maximal response. After recording 20 min of stable baseline, the stimulation was stopped and 20 μ M ketamine was applied for 30 min, then stimulation was resumed. Field potentials were filtered at 2 kHz, acquired and digitized at 10 kHz on a personal computer using custom software (LabVIEW, National Instruments). Synaptic strength was measured as the initial slope (10–40% of the rising phase) of the field potential. The group data were analysed as follows: (1) the initial slopes of the field potential were expressed as percentages of the preconditioning baseline average; (2) the timescale in each experiment was converted to the time from the end of ketamine application; and (3) the time-matched, normalized data were averaged across experiments.

30. Duman, C. H., Schlesinger, L., Kodama, M., Russell, D. S. & Duman, R. S. A role for MAP kinase signalling in behavioral models of depression and antidepressant treatment. *Biol. Psychiatry* **61**, 661–670 (2007).
31. Porsolt, R. D., Le Pichon, M. & Jalfre, M. Depression: A new animal model sensitive to antidepressant treatments. *Nature* **266**, 730–732 (1977).
32. Park, S. *et al.* Elongation factor 2 and fragile X mental retardation protein control the dynamic translation of Arc/Arg3.1 essential for mGluR-LTD. *Neuron* **59**, 70–83 (2008).
33. Kavalali, E. T., Klingauf, J. & Tsien, R. W. Activity-dependent regulation of synaptic clustering in a hippocampal culture system. *Proc. Natl Acad. Sci. USA* **96**, 12893–12900 (1999).

Structure-based design of non-natural amino-acid inhibitors of amyloid fibril formation

Stuart A. Sievers^{1*}, John Karanicolas^{2,3*}, Howard W. Chang^{1*}, Anni Zhao^{1*}, Lin Jiang^{1*}, Onofrio Zirafi⁴, Jason T. Stevens³, Jan Münch⁴, David Baker² & David Eisenberg¹

Many globular and natively disordered proteins can convert into amyloid fibrils. These fibrils are associated with numerous pathologies¹ as well as with normal cellular functions^{2,3}, and frequently form during protein denaturation^{4,5}. Inhibitors of pathological amyloid fibril formation could be useful in the development of therapeutics, provided that the inhibitors were specific enough to avoid interfering with normal processes. Here we show that computer-aided, structure-based design can yield highly specific peptide inhibitors of amyloid formation. Using known atomic structures of segments of amyloid fibrils as templates, we have designed and characterized an all-D-amino-acid inhibitor of the fibril formation of the tau protein associated with Alzheimer's disease, and a non-natural L-amino-acid inhibitor of an amyloid fibril that enhances sexual transmission of human immunodeficiency virus. Our results indicate that peptides from structure-based designs can disrupt the fibril formation of full-length proteins, including those, such as tau protein, that lack fully ordered native structures. Because the inhibiting peptides have been designed on structures of dual- β -sheet 'steric zippers', the successful inhibition of amyloid fibril formation strengthens the hypothesis that amyloid spines contain steric zippers.

The finding that dozens of pathologies, including Alzheimer's disease, are associated with amyloid fibrils has stimulated research on fibril inhibition. One approach uses the self-associating property of proteins that form fibrils to poison fibril formation with short peptide segments^{6–11}. A second approach is based on screening for molecules that can disrupt fibril formation^{12,13}. Here we take a third approach to fibril inhibition: structure-based design of non-natural peptides targeted to block the ends of fibrils. With advanced sampling techniques and by minimizing an appropriate energy function, we identify novel candidate inhibitors computationally from a large peptide space that interact favourably with our template structure. This approach has been made possible by the determination of several dozen fibril-like atomic structures of segments from amyloid-forming proteins^{14–16}.

These structures reveal a common motif called a steric zipper, in which a pair of β -sheets is held together by the interdigitation of their side chains¹⁴. Using as templates the steric-zipper structures formed by segments of two pathological proteins, we have designed inhibitors that cap fibril ends. As we show, the inhibitors greatly slow the fibril formation of the parent proteins of the segments, offering a route to designed chemical interventions and supporting the hypothesis that steric zippers are the principal structural elements of these fibrils.

One of the two fibril-like steric zippers that we have chosen as a target for inhibitor design is the hexapeptide VQIVYK, residues 306–311 of the tau protein, which forms intracellular amyloid fibrils in Alzheimer's disease¹⁷. This segment has been shown to be important for fibril formation of the full-length protein and itself forms fibrils

with biophysical properties similar to full-length tau fibrils^{15,18,19}. Our second template for inhibitor design, identified by the '3D profile' algorithm^{20,21}, is the steric-zipper structure of the peptide segment GGVLVN from the amyloid fibril formed by ²⁴⁸PAP²⁸⁶, a proteolytic fragment containing residues 248–286 of prostatic acid phosphatase, a protein abundant in semen. ²⁴⁸PAP²⁸⁶ fibrils, also known as semen-derived enhancer of virus infection (SEVI), enhance human immunodeficiency virus (HIV) infection by orders of magnitude in cell culture studies, whereas the monomeric peptide is inactive²².

Our computational approach to designing non-natural peptides that inhibit fibril formation is summarized in Fig. 1 for the VQIVYK segment of tau protein; the same general strategy is used for the GGVLVN segment of ²⁴⁸PAP²⁸⁶. In both systems, we design a tight interface between the inhibiting peptide and the end of the steric zipper to block additional segments from joining the fibril. By sampling L or D amino acids, or commercially available non-natural amino acids, we can design candidate inhibitors with side chains that maximize hydrogen bonding and hydrophobic interactions across the interface.

We propose that the steric-zipper structures of the VQIVYK and GGVLVN segments represent the spines of the fibrils formed by the parent proteins containing these segments. Supporting our hypothesis are our results that D-amino-acid inhibitors designed on the VQIVYK steric-zipper template inhibit fibril formation not only of the VQIVYK segment, but also of two tau constructs, K12 and K19^{23,24} (Fig. 2a). Similarly, the peptide composed of non-natural amino acids designed on the GGVLVN template inhibits the fibril formation of ²⁴⁸PAP²⁸⁶ and greatly inhibits the HIV infectivity of human cells in culture.

To design a D-amino-acid hexapeptide sequence that interacts favourably with the VQIVYK steric zipper¹⁵, and prevents further addition of tau molecules to the fibril, we used the Rosetta software²⁵. This led to the identification of four D-amino-acid peptides: D-TLKIVW, D-TWKLVL, D-DYYFEF and D-YVIER, in which the prefix signifies that all α -carbon atoms are in the D configuration (Fig. 2b, c, Supplementary Figs 1 and 2 and Supplementary Table 1). In the D-TLKIVW design model (Fig. 2b, c and Supplementary Fig. 1), the inhibitor packs tightly across the top of the VQIVYK steric-zipper structure, maintaining all main-chain hydrogen bonds. The side-chain hydrogen bonding between layers of stacked Gln 307 residues is replaced in the designed interface by an interaction with D-Lys3. Several hydrophobic interactions between D-TLKIVW and the two VQIVYK β -strands contribute to the favourable binding energy (Supplementary Table 1). In the design, the D-peptide blocks the addition of another layer of VQIVYK, both above the D-peptide and across on the mating β -sheet (Supplementary Fig. 3). D-Leu 2 of the designed inhibitor prevents the addition of a VQIVYK molecule above it through a steric clash with Ile 308 of VQIVYK and on the mating sheet through a clash with Val 306 and Ile 308 (Supplementary Fig. 3). These steric clashes involving D-Leu 2 are intended to block fibril growth.

¹Departments of Biological Chemistry and Chemistry and Biochemistry, Howard Hughes Medical Institute, UCLA, Box 951970, Los Angeles, California 90095-1570, USA. ²Department of Biochemistry and Howard Hughes Medical Institute, University of Washington, Seattle, Washington 98195, USA. ³Center for Bioinformatics and Department of Molecular Biosciences, University of Kansas, 1200 Sunnyside Avenue, Lawrence, Kansas 66045-7534, USA. ⁴Institute of Molecular Virology, University Hospital Ulm, Meyerhofstrasse 1, 89081 Ulm, Germany.

*These authors contributed equally to this work.

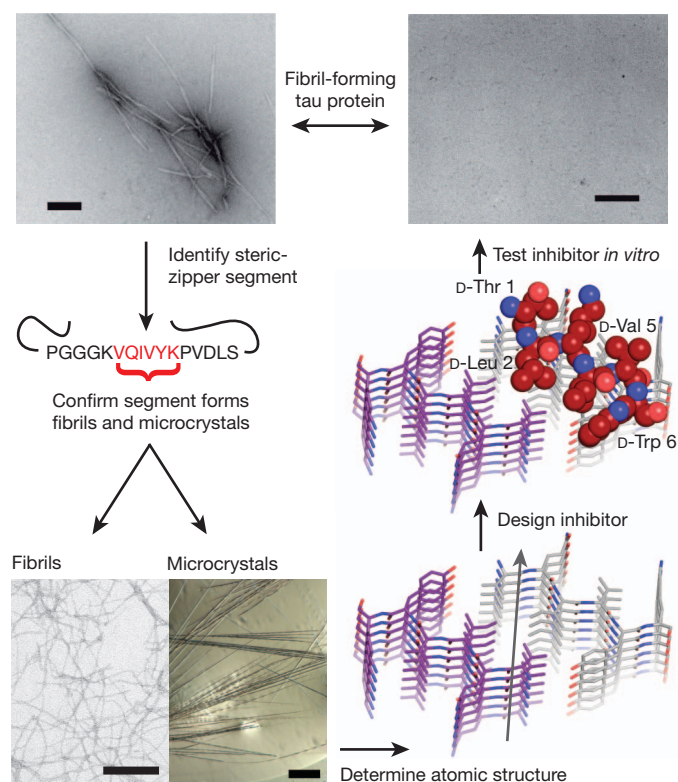


Figure 1 | Design and characterization of peptide inhibitors of amyloid fibril formation. Tau constructs form fibrils *in vitro*²⁴ (top left; scale bar, 200 nm). The VQIVYK segment in isolation forms fibrils and microcrystals (bottom left; fibril scale bar, 200 nm; microcrystal scale bar, 100 μ m). The atomic structure of the fibril-like VQIVYK segment reveals a characteristic steric-zipper motif¹⁵ comprising a pair of interacting β -sheets (purple and grey) running along the fibril axis (grey arrow) (bottom right). We designed a D-amino-acid peptide to bind to the end of the steric-zipper template and prevent fibril elongation (middle right). The D-peptide (red) is designed to satisfy hydrogen bonds and make favourable non-polar interactions with the molecule below, while preventing the addition of other molecules above and on the opposite β -sheet. As shown *in vitro*, the designed D-peptide prevents the formation of fibrils when incubated with tau K19 (upper right; scale bar, 200 nm).

We used fluorescence spectroscopy and electron microscopy to assess whether the designed D-peptides inhibit the fibril formation of the tau segment VQIVYK and of the tau constructs K12 and K19. Of our designed inhibitors, D-TLKIVW is the most effective (Supplementary Fig. 4). Electron microscopy, performed after three days, verified that incubation with equimolar D-TLKIVW prevents K19 fibril formation, which would otherwise have occurred within the elapsed time (Fig. 1, upper right). D-TLKIVW delays fibril formation of VQIVYK, K12 and K19 even when present in sub-equimolar concentration (Supplementary Fig. 5). A fivefold molar excess of D-TLKIVW delays K12 fibril formation for more than two weeks in some experimental replicates (Supplementary Fig. 5c, d). In tenfold molar excess, D-TLKIVW prevents the fibril formation of K12 for more than 60 hours in the presence of preformed K12 fibril seeds, suggesting that the peptide interacts with fibrils (Fig. 2d). Also, kinetic analysis shows that the fibril elongation rate decreases in the presence of increasing concentrations of inhibitor peptide (Supplementary Fig. 6). The large increase in lag time in unseeded reactions may be due to interactions with small aggregates formed during the process of fibril formation.

To investigate the specificity of the designed inhibitor, we tested scrambled sequence variants of D-TLKIVW that have poor (that is, high) calculated energies and unfavourable packing (Supplementary Table 1). The scrambled peptides D-TIKWVL, D-TIWKVL and D-LKTWIV have little inhibitory effect when present at an equimolar

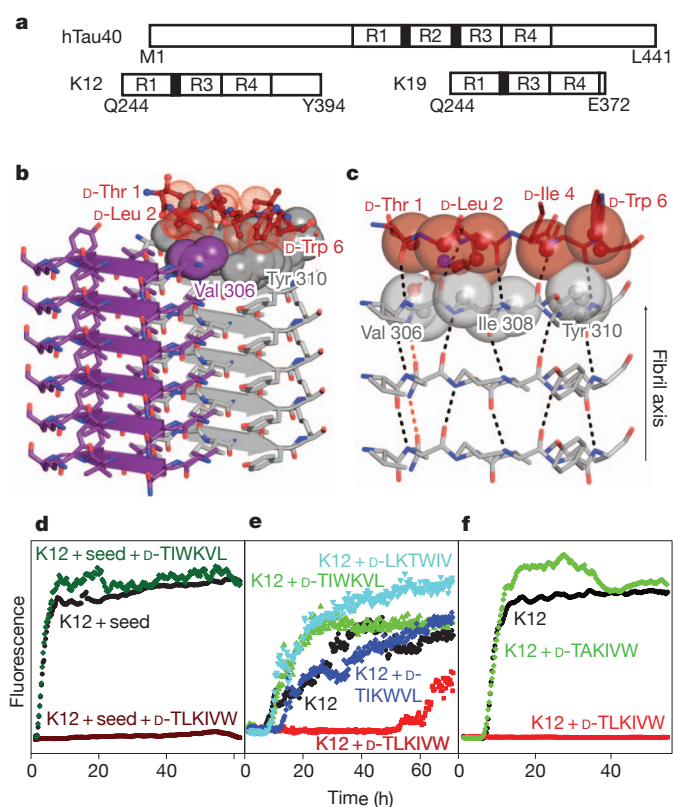


Figure 2 | Designed D-peptide delays tau K12 fibril formation in a sequence-specific manner. **a**, Tau construct composition²³. The longest human tau isoform found in the central nervous system, hTau40 (Uniprot ID, P10636-8), contains four microtubule-binding repeats, R1 to R4, whereas K12 and K19 lack R2. The black bars at the amino termini of R2 and R3 represent the fibrillogenic segments VQIINK and VQIVYK, respectively. **b**, The inhibitor D-TLKIVW (red) is designed to interact with atoms on both β -strands of the VQIVYK steric zipper (grey) primarily through hydrophobic packing and hydrogen-bonding interactions. **c**, The inhibitor interacts with the VQIVYK β -strand below. The transparent spheres show where the two molecules interact favourably. Black and red dashes indicate main-chain and side-chain hydrogen bonds, respectively. Stereo views of **b** and **c** are shown in Supplementary Fig. 1. **d**, The seeded fibril formation of 50 μ M K12 in the presence and absence of a tenfold molar excess of peptide was monitored by Thioflavin S fluorescence. In the presence of the scrambled peptide D-TIWKVL (dark green) and alone (black), seeded K12 fibril formation occurs with almost no lag time. However, D-TLKIVW prevents fibril formation for days (maroon). **e**, At equimolar concentrations, D-TLKIVW (red) inhibits the fibril formation of 50 μ M K12. D-TIKWVL (blue), with only three residues scrambled, shows weak inhibition. However, no inhibition is observed for either D-TIWKVL (green) or D-LKTWIV (cyan). **f**, The replacement of D-Leu 2, designed to clash with VQIVYK on the opposite sheet, with D-Ala eliminates the inhibition of fibril formation.

ratio with VQIVYK, K12 and K19 (Fig. 2e and Supplementary Fig. 7), showing that the inhibition is sequence specific. Also, the diastereomer, L-TLKIVW, is less effective than D-TLKIVW (Supplementary Fig. 8). As a further test of the specificity of our design, we confirmed that D-TLKIVW is unable to block the fibril formation of amyloid- β , which also is associated with Alzheimer's disease (Supplementary Fig. 9). This suggests that the D-peptide inhibitor is not general to amyloid systems, but is specific to the VQIVYK interface in tau protein. Such specificity is essential for designed inhibitors if they are not to interfere with proteins that natively function in an amyloid state³.

To confirm that the designed D-peptide inhibits in accordance with the design model (Fig. 2b, c and Supplementary Fig. 1), we performed several additional tests. First we visualized the position of the inhibitor D-TLKIVW relative to fibrils of the tau construct K19 using electron microscopy. We covalently linked Monomaleimido Nanogold particles

both to the inhibitor and, separately, to a scrambled hexapeptide, D-LKTWIV. We used a blind counting assay and found that, relative to Nanogold alone, D-TLKIVW shows a significant binding preference for the end of fibrils, in contrast to the scrambled control peptide, D-LKTWIV (Fig. 3a and Supplementary Fig. 10).

As a further test of the model, we used NMR to characterize the binding affinity of D-TLKIVW for tau fibrils. The ^1H NMR spectra for D-TLKIVW were collected in the presence of increasing concentrations of VQIVYK or K19 fibrils. Because neither K19 nor VQIVYK contains tryptophan, we were able to monitor the ^1H resonance of the indole proton of the tryptophan in our inhibitor. When bound to a fibril, the inhibitor, D-TLKIVW, is removed from the soluble phase and the ^1H resonance is diminished²⁶ (Fig. 3b and Supplementary Fig. 11). As a control, we also measured spectra for the non-inhibiting peptide D-LKTWIV present with D-TLKIVW in the same reaction mixture. As shown in Fig. 3b, the presence of VQIVYK fibrils at a given concentration reduces the D-TLKIVW indole resonance much more than it does the D-LKTWIV indole resonance. Spectra of the two peptides are shown in Supplementary Fig. 12. By monitoring the D-TLKIVW indole resonance over a range of VQIVYK fibril concentrations, we estimate the apparent dissociation constant of the interaction between D-TLKIVW and VQIVYK fibrils to be $\sim 2\ \mu\text{M}$ (Supplementary Fig. 11a and Methods). This value corresponds to a standard free binding energy of $\sim 7.4\ \text{kcal mol}^{-1}$, with $\sim 2.5\ \text{kcal mol}^{-1}$ from non-polar interactions and $\sim 4.9\ \text{kcal mol}^{-1}$ from six hydrogen bonds (Methods). Repeating the NMR binding experiment with K19 fibrils yields a similar trend (Supplementary Fig. 11b). To determine whether D-TLKIVW has affinity for soluble VQIVYK, we measured ^1H NMR spectra of D-TLKIVW and D-LKTWIV in the presence of increasing amounts of soluble VQIVYK. Only a slight change in the respective chemical shifts of the indole proton peaks of D-TLKIVW and D-LKTWIV is observed, even at a 70-fold molar excess of VQIVYK (Supplementary Fig. 13). This, together with the ability of the peptide to prevent seeded fibril formation, suggests that D-TLKIVW does not interact with monomers but rather with a structured, fibril-like species.

As another test of our design model, we replaced the D-Leu residue with D-Ala in D-TLKIVW. Our structural model suggests that D-Leu 2 of D-TLKIVW is important for preventing tau fibril formation because of its favourable interaction with the Ile residue of the VQIVYK molecule below and with Ile and the first Val of VQIVYK across the steric zipper (Fig. 2b, c and Supplementary Fig. 1). The D-Ala replacement

eliminates these interactions and, furthermore, removes a steric clash that would occur were another VQIVYK molecule placed across from the inhibitor (Supplementary Fig. 3 and Supplementary Table 1). When the D-Ala variant is incubated with VQIVYK and the tau constructs, it has no inhibitory effect on fibril formation (Fig. 2f and Supplementary Fig. 14). This confirms that D-Leu 2 is critical for the efficacy of D-TLKIVW, consistent with our model.

In summary, although our electron microscopy, NMR and D-Ala replacement results support a model in which the designed peptide D-TLKIVW binds to the ends of tau fibrils, they do not constitute proof that the inhibitors bind exactly as anticipated in the designs (Supplementary Fig. 15).

To expand on our design methodology, we computationally designed an inhibitor of $^{248}\text{PAP}^{286}$ fibril formation containing non-natural L-amino acids (Fig. 4b and Supplementary Fig. 16), using the GGVLVN structure as a template (Fig. 4a and Supplementary Table 2). This peptide, Trp-His-Lys-chAla-Trp-hydroxyTic (WW61), contains an Ala derivative, β -cyclohexyl-L-alanine (chAla) and a Tyr/Pro derivative, 7-hydroxy-(S)-1,2,3,4-tetrahydroisoquinoline-3-carboxylic acid (hydroxyTic), both of which increase contact area with the GGVLVN template. The non-natural chAla forms hydrophobic interactions with the Leu residue in the steric-zipper interface, and hydroxyTic supports the favourable placement of chAla through hydrophobic packing (Fig. 4b and Supplementary Fig. 16b).

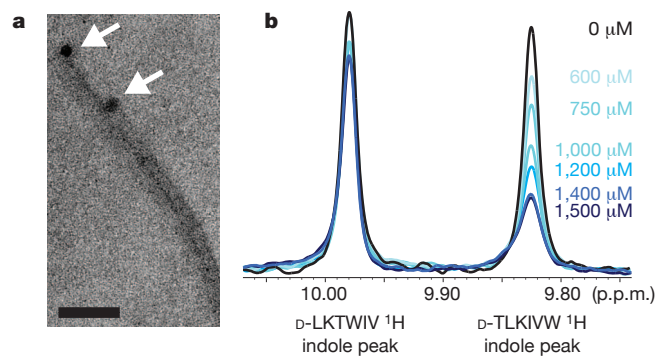


Figure 3 | Mechanism of interaction. **a**, Nanogold covalently bound to D-TLKIVW localizes at the ends (arrows) of two tau K19 fibrils. Scale bar, 50 nm. **b**, The inhibitor D-TLKIVW binds to fibrils with an estimated affinity constant in the low micromolar range, as shown by the indole proton region of the 500-MHz ^1H NMR spectra of D-TLKIVW (9.83 p.p.m.) and D-LKTWIV (9.98 p.p.m.) in the presence of increasing concentrations of VQIVYK fibrils. The resonance of the D-TLKIVW indole proton is reduced in the presence of increasing concentrations of VQIVYK fibrils, whereas the indole proton signal for the scrambled control peptide D-LKTWIV is only slightly affected. Fibril solutions contained 0–1,500 μM VQIVYK monomers, as indicated.

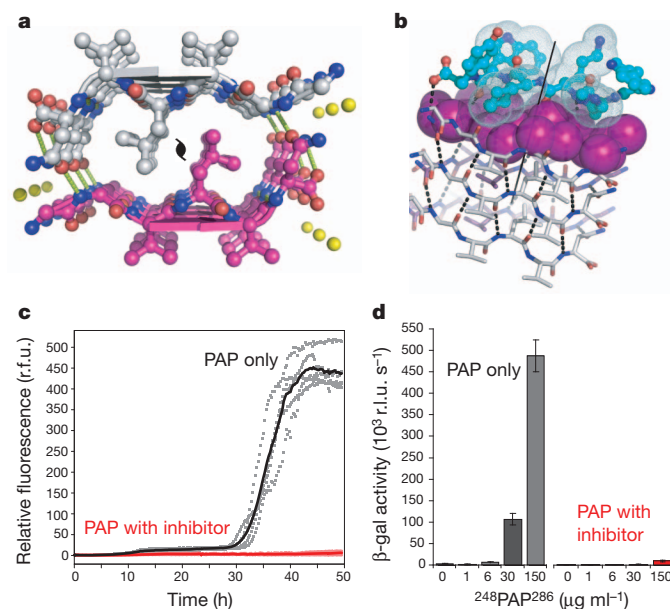


Figure 4 | Designed non-natural peptide inhibits $^{248}\text{PAP}^{286}$ fibril formation. **a**, The view down the fibril axis of the crystal structure of the GGVLVN steric zipper reveals two mating β -sheets with parallel, in-register β -strands (hydrogen bonds, green dashed lines; water molecules, yellow spheres). **b**, View roughly perpendicular to a fibril of three layers, with the atoms of the side chains of the top layer shown as purple spheres. On top is a designed non-natural peptide inhibitor, Trp-His-Lys-chAla-Trp-hydroxyTic (blue; see Supplementary Fig. 16). **c**, The inhibitor blocks $^{248}\text{PAP}^{286}$ fibril formation, as shown by monitoring Thioflavin T fluorescence. With a twofold molar excess of the inhibitor (pale red), the fluorescence remains low over the course of the experiment for all five replicates, unlike in the absence of inhibitor (grey). Mean fluorescence values are shown as solid red and black lines with and without the inhibitor, respectively. r.f.u., relative fluorescence units. **d**, HIV infection rates were determined by monitoring β -galactosidase (β -gal) activity. Agitated $^{248}\text{PAP}^{286}$ alone efficiently increases viral infection, whereas $^{248}\text{PAP}^{286}$ mixtures incubated with inhibitor were unable to enhance HIV infection. Peptide concentrations during virion treatment are indicated on the x axis. Error bars show the s.d. of three measurements per sample. r.l.u., relative light units.

Moreover, we propose that the bulky side chains and steric constraints of hydroxyTic provide hindrance to further fibril growth.

This designed peptide, WW61, effectively delays both seeded and unseeded fibril formation of $^{248}\text{PAP}^{286}$ *in vitro* (Fig. 4c and Supplementary Figs 17 and 18). In the presence of a twofold molar excess of this inhibitor, seeded fibril formation is efficiently blocked for more than two days (Fig. 4c). Furthermore, we see that increasing the concentration of this inhibitor extends the fibril formation lag time (Supplementary Fig. 19). These inhibition assay results were further confirmed by electron microscopy (Supplementary Fig. 20). As a control for specificity, we tested the effect of GIHKQK, from the amino terminus of $^{248}\text{PAP}^{286}$, and PYKLWN, a peptide with the same charge as WW61. Neither peptide affected fibril formation kinetics, indicating that the inhibitory activity of the designed peptide is sequence specific (Supplementary Fig. 21).

Because $^{248}\text{PAP}^{286}$ fibrils (SEVI) have been shown to enhance HIV infection²², using a functional assay we investigated whether WW61 is able to prevent this enhancement. In this experiment, we treated HIV particles with $^{248}\text{PAP}^{286}$ solutions that had been agitated for 20 hours (to allow fibril formation) in the presence or absence of WW61, and infected TZM-bl indicator cells. As has been previously observed, SEVI efficiently enhanced HIV infection²². However, $^{248}\text{PAP}^{286}$ incubated with the designed inhibitor prevented HIV infection (Fig. 4d).

We performed several control experiments to verify that the lack of infectivity observed in the assay is indeed due to the inhibition of SEVI formation. First we confirmed that in the absence of SEVI the designed inhibitor WW61 does not affect HIV infectivity (Supplementary Fig. 22a). We also found that the control peptides GIHKQK and PYKLWN, which do not inhibit $^{248}\text{PAP}^{286}$ fibril formation, fail to decrease HIV infectivity (Supplementary Fig. 22b). Additionally, we observed that WW61 has no inhibitory effect on polylysine-mediated HIV infectivity²⁷, further ruling out a non-specific electrostatic interaction mechanism (Supplementary Fig. 22a). Together, these results demonstrate that a peptide capable of preventing $^{248}\text{PAP}^{286}$ fibril formation also inhibits the generation of virus-enhancing material.

Structure-based design of inhibitors of amyloid fibril formation has been challenging in the absence of detailed information about the atomic-level interactions that form the fibril spine. So far, one of the most successful structure-based approaches to preventing fibril formation has been to stabilize the native tetrameric structure of transthyretin²⁸. That approach is well suited to the prevention of fibril formation of proteins with known native structures, but other proteins involved in amyloid-related diseases, such as tau protein, amyloid- β and $^{248}\text{PAP}^{286}$, lack fully ordered native structures²⁹. Our structure-based approach makes it possible to design inhibitors independent of native structure. Instead, the templates are atomic-level structures of short, fibril-forming segments^{14,15}. By using these fibril-like templates, and adopting computational methods successful in designing novel proteins and protein-protein interfaces^{25,30}, we have created specific inhibitors of proteins that normally form fibrils. These results support the hypothesis that the steric zipper is a principal feature of tau-related and SEVI fibrils, and suggest that, with current computational methods and steric-zipper structures, we have the tools to design specific inhibitors to prevent the formation of other amyloid fibrils.

METHODS SUMMARY

We used crystal structures of hexapeptide segments of VQIVYK and GGVLVN as templates to design peptide inhibitors using the Rosetta software²⁵. Briefly, this algorithm searches possible side-chain conformations (called rotamers) of all amino acids in a peptide β -strand backbone stacked onto the fibril end of both segment structures. The Rosetta software is extended to sample the approximate side-chain conformation of non-natural D and L amino acids by adapting side-chain torsion angles from those in their natural counterparts. The lowest energy set of side-chain rotamers is identified by combinatorial optimization of a potential consisting of a term for the Lennard-Jones potential, an orientation-dependent hydrogen-bond potential term, an implicit solvation term and a structure-derived side-chain and backbone torsional potential term.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 6 December 2010; accepted 21 April 2011.

Published online 15 June 2011.

1. Westermark, P. *et al.* A primer of amyloid nomenclature. *Amyloid* **14**, 179–183 (2007).
2. Maji, S. K. *et al.* Functional amyloids as natural storage of peptide hormones in pituitary secretory granules. *Science* **325**, 328–332 (2009).
3. Fowler, D. M., Koulou, A. V., Balch, W. E. & Kelly, J. W. Functional amyloid – from bacteria to humans. *Trends Biochem. Sci.* **32**, 217–224 (2007).
4. Astbury, W. T. & Dickinson, S. The X-ray interpretation of denaturation and the structure of the seed globulins. *Biochem. J.* **29**, 2351–2360 (1935).
5. Calamai, M., Chiti, F. & Dobson, C. M. Amyloid fibril formation can proceed from different conformations of a partially unfolded protein. *Biophys. J.* **89**, 4201–4210 (2005).
6. Tjernberg, L. O. *et al.* Arrest of β -amyloid fibril formation by a pentapeptide ligand. *J. Biol. Chem.* **271**, 8545–8548 (1996).
7. Findeis, M. A. Peptide inhibitors of β amyloid aggregation. *Curr. Top. Med. Chem.* **2**, 417–423 (2002).
8. Sciarretta, K. L., Gordon, D. J. & Meredith, S. C. Peptide-based inhibitors of amyloid assembly. *Methods Enzymol.* **413**, 273–312 (2006).
9. Soto, C., Kindy, M. S., Baumann, M. & Frangione, B. Inhibition of Alzheimer's amyloidosis by peptides that prevent β -sheet conformation. *Biochem. Biophys. Res. Commun.* **226**, 672–680 (1996).
10. Kokkoni, N., Stott, K., Amijee, H., Mason, J. M. & Doig, A. J. N-methylated peptide inhibitors of amyloid aggregation and toxicity. Optimization of the inhibitor structure. *Biochemistry* **45**, 9906–9918 (2006).
11. Sato, T. *et al.* Inhibitors of amyloid toxicity based on β -sheet packing of A β 40 and A β 42. *Biochemistry* **45**, 5503–5516 (2006).
12. Larbig, G., Pickhardt, M., Lloyd, D. G., Schmidt, B. & Mandelkow, E. Screening for inhibitors of tau protein aggregation into Alzheimer paired helical filaments: a ligand based approach results in successful scaffold hopping. *Curr. Alzheimer Res.* **4**, 315–323 (2007).
13. Wiesehan, K. *et al.* Selection of D-amino-acid peptides that bind to Alzheimer's disease amyloid peptide $\text{A}\beta_{1-42}$ by mirror image phage display. *ChemBioChem* **4**, 748–753 (2003).
14. Nelson, R. *et al.* Structure of the cross- β spine of amyloid-like fibrils. *Nature* **435**, 773–778 (2005).
15. Sawaya, M. R. *et al.* Atomic structures of amyloid cross- β spines reveal varied steric zippers. *Nature* **447**, 453–457 (2007).
16. Wiltzius, J. J. *et al.* Molecular mechanisms for protein-encoded inheritance. *Nature Struct. Mol. Biol.* **16**, 973–978 (2009).
17. Selkoe, D. J. Alzheimer's disease: genes, proteins, and therapy. *Physiol. Rev.* **81**, 741–766 (2001).
18. Goux, W. J. *et al.* The formation of straight and twisted filaments from short tau peptides. *J. Biol. Chem.* **279**, 26868–26875 (2004).
19. von Bergen, M. *et al.* Assembly of τ protein into Alzheimer paired helical filaments depends on a local sequence motif ($^{306}\text{VQIVYK}^{311}$) forming β structure. *Proc. Natl Acad. Sci. USA* **97**, 5129–5134 (2000).
20. Goldschmidt, L., Teng, P. K., Riek, R. & Eisenberg, D. Identifying the amyloids, proteins capable of forming amyloid-like fibrils. *Proc. Natl Acad. Sci. USA* **107**, 3487–3492 (2010).
21. Thompson, M. J. *et al.* The 3D profile method for identifying fibril-forming segments of proteins. *Proc. Natl Acad. Sci. USA* **103**, 4074–4078 (2006).
22. Münch, J. *et al.* Semen-derived amyloid fibrils drastically enhance HIV infection. *Cell* **131**, 1059–1071 (2007).
23. Friedhoff, P., von Bergen, M., Mandelkow, E. M., Davies, P. & Mandelkow, E. A nucleated assembly mechanism of Alzheimer paired helical filaments. *Proc. Natl Acad. Sci. USA* **95**, 15712–15717 (1998).
24. Wille, H., Drewes, G., Biernat, J., Mandelkow, E. M. & Mandelkow, E. Alzheimer-like paired helical filaments and antiparallel dimers formed from microtubule-associated protein tau *in vitro*. *J. Cell Biol.* **118**, 573–584 (1992).
25. Kuhlman, B. *et al.* Design of a novel globular protein fold with atomic-level accuracy. *Science* **302**, 1364–1368 (2003).
26. Chen, Z., Krause, G. & Reif, B. Structure and orientation of peptide inhibitors bound to β -amyloid fibrils. *J. Mol. Biol.* **354**, 760–776 (2005).
27. Roan, N. R. *et al.* The cationic properties of SEVI underlie its ability to enhance human immunodeficiency virus infection. *J. Virol.* **83**, 73–80 (2009).
28. Petrassi, H. M., Klabunde, T., Sacchettini, J. & Kelly, J. W. Structure-based design of N-phenyl phenoxazine transthyretin amyloid fibril inhibitors. *J. Am. Chem. Soc.* **122**, 2178–2192 (2000).
29. Schweers, O., Schonbrunn-Hanebeck, E., Marx, A. & Mandelkow, E. Structural studies of tau protein and Alzheimer paired helical filaments show no evidence for β -structure. *J. Biol. Chem.* **269**, 24290–24297 (1994).
30. Fleishman, S. J. *et al.* Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. *Science* **332**, 816–821 (2011).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. I. Ivanova, J. Corn, T. Kortemme, D. Anderson, M. R. Sawaya, M. Phillips, S. Sambashivan, J. Park, M. Landau, A. Laganowsky, Q. Zhang, R. Clubb, F. Guo, T. Yeates, J. Nowick, J. Zheng and M. J. Thompson for discussions; the HHMI, NIH, NSF, Gates Foundation and Joint Center for Translational Medicine for support; R. Peterson for help with NMR experiments; E. Mandelkow for providing tau constructs; R. Riek for providing amyloid- β ; and J. Stroud for amyloid- β preparation.

Support came from the Damon Runyon Cancer Research Foundation (J.K.), the Ruth L. Kirschstein National Research Service Award (H.W.C.), the programme for junior professors by the Ministry of Science, Baden-Württemberg (J.M.), and a UCLA-IGERT bioinformatics traineeship (S.A.S.).

Author Contributions S.A.S., J.K., D.B., J.M. and D.E. designed the project. J.K. and S.A.S. created the design protocol. J.K. designed the D-peptides. L.J. expanded the design methodology and designed the non-natural amino-acid peptides. S.A.S., H.W.C. and A.Z. performed the fluorescence experiments and electron microscopy, and analysed kinetic data. A.Z. determined the structure of GGVLVN. O.Z. performed the HIV

infectivity experiments. J.T.S. determined the tau fibril elongation rates. S.A.S. performed the NMR experiments. S.A.S., J.K. and D.E. wrote the manuscript and coordinated contributions by other authors.

Author Information Atomic coordinates and structure factors for the reported GGVLVN structure have been deposited in the Protein Data Bank with accession code 3PPD. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to D.E. (david@mbi.ucla.edu).

METHODS

Computational design. Computational designs were carried out using the Rosetta software²⁵ (<http://www.rosettacommons.org>). This algorithm involves building side-chain rotamers of all amino acids onto a fixed protein backbone. The lowest energy set of side-chain rotamers is then identified as those which minimize an energy function containing a Lennard–Jones potential, an orientation-dependent hydrogen-bond potential, a solvation term, amino-acid-dependent reference energies and a statistical torsional potential that depends on the backbone and side-chain dihedral angles.

D-amino-acid tau inhibitors. The crystal structure of VQIVYK (ref. 15; Protein Data Bank ID, 2ON9) was used as a starting scaffold for computational design. To take full advantage of the statistical nature of the rotamer library and some terms in the Rosetta energy function, the stereochemistry of the fibril scaffold was inverted so that design would take place using L amino acids. An extended L-peptide was aligned with the N, C and O backbone atoms of the D-fibril scaffold. This L-peptide was subsequently redesigned, keeping all atoms of the D-fibril fixed. The stereochemistry of the final design model was then inverted, yielding a D-peptide designed to cap an L-fibril. We inspected the finished models to confirm that inversion of the stereochemistry at the Thr and Ile C β atoms did not make the designs energetically unfavourable. Energetic consequences of incorporating a D inhibitor peptide in the middle of an L fibril were subsequently evaluated to ensure that fibril propagation could not continue after association of an inhibitor. Calculations of the area buried and shape complementarity were performed with AREAIMOL³¹ and SC³², respectively.

L-peptide²⁴⁸PAP²⁸⁶ inhibitors. The crystal structure of GGVLVN (PDB ID, 3PPD) was used as a template for the following design procedure. An extended L-peptide was aligned according to crystal symmetry. Small, random perturbations of the L-peptide were performed to optimize the rigid-body arrangement between the fibril template and the peptide inhibitor. Full sequence optimization of the inhibitor was performed using the Rosetta software package, allowing residues directly contacting the inhibitor to repack; other scaffold residues remained fixed. Because the design calculations use a discrete rotamer representation of the side chains, we next performed simultaneous quasi-Newtonian optimization of the inhibitor rigid-body orientation, the side-chain torsion angles and, in some cases, the backbone torsion angles using the full-atom Rosetta energy function. This optimization was essential to the subsequent assessment of the inhibition of the design. Several iterative runs of small perturbations in inhibitor placement, interface design and refinement were performed to improve hydrogen-bonding and packing interactions. The designs that ranked highest on the basis of the total binding energy between the inhibitor and the fibril scaffold and the interfacial shape complementarity³² were subsequently synthesized and tested.

For each initial active L-peptide design, the non-natural L amino acids were incorporated using a growth strategy. Non-natural amino acids, structurally similar to those of initial active designs, were selected on the basis of their solubility, side-chain shape and commercial availability. Side-chain conformations were approximately sampled by adopting side-chain torsion angles from those in their natural counterparts. Sequence optimization of the inhibitor was performed and the optimal set of rotamers identified using Monte Carlo simulated annealing with the full-atom energy function described above. The resulting designs were ranked on the basis of the total binding energy between the inhibitor and the fibril scaffold. **Tau construct expression and purification.** pNG2 expression vectors (derived from pET-3b³³) containing either the K12 or K19 gene were provided by E. Mandelkow³⁴. Expression in BL21(DE3) *Escherichia coli*³³ was induced with 1 mM isopropyl thiogalactoside when the absorbance $A_{600\text{ nm}}$ was between 0.8 and 1.0, and cells were collected after 3–4 h. K12 and K19 were purified on the basis of previously described methods³⁵. Cells were pelleted for 20 min at 4,700g and resuspended in 20 mM MES, pH 6.8, 1 mM EDTA, 0.2 mM MgCl₂, 5 mM DTT, 1 mM PMSF and a protease inhibitor cocktail. The cells were sonicated for 2.5 min and, following addition of NaCl to bring cell lysate to 0.5 M NaCl, the lysate was boiled for 20 min. The lysate was sedimented at 30,000g for 20 min and dialyzed twice against 20 mM MES, pH 6.8, 50 mM NaCl, 1 mM EDTA, 1 mM MgCl₂, 2 mM DTT and 0.1 mM PMSF at 4 °C. The dialysate was pelleted for 20 min at 30,000g and filtered before cation exchange chromatography on an AKTA Explorer (GE Pharmacia) with a HighTrap HP SP 5-ml column (GE Healthcare). The sample was eluted with a linear gradient of up to 60% buffer B (20 mM MES, pH 6.8, 1 M NaCl, 1 mM EDTA, 1 mM MgCl₂, 2 mM DTT and 0.1 mM PMSF). Size exclusion chromatography was optionally performed with a Superdex 75 10/300 GL column (GE Healthcare) in PBS buffer (137 mM NaCl, 3 mM KCl, 10 mM Na₂HPO₄, 2 mM KH₂PO₄, pH 7.4) with 1 mM DTT on the AKTA Explorer depending on preparation purity as assessed by SDS polyacrylamide gel electrophoresis.

Tau construct inhibition assays. Fibril formation assays were performed on the basis of previously published protocols^{35–38}. Reaction mixtures (150 μ l) containing

50 μ M tau K12 or K19, as determined by the Micro BCA Protein Assay Kit (Pierce), were incubated in 250 mM sodium phosphate buffer, pH 7.4, with 1 mM DTT, 12.5 μ M heparin (average molecular mass, 6,000 Da; Sigma) and 10 μ M Thioflavin S (ThS; MP Bio). Inhibitor peptides (CS Bio, Celtek Biosciences) were dissolved in 250 mM phosphate buffer, pH 7.4, to 0.5 mM and added at specified molar ratios. Reactions were split into a minimum of three replicates in black, 96-well, optically clear plates (Nunc), sealed with Corning pressure-sensitive sealing tape and monitored using either a Varioskan plate reader (Thermo Scientific), for K12, or a SpectraMax M5, for K19. The fluorescence signal was measured every 15 min with excitation and emission wavelengths of 440 and 510 nm, respectively, at 37 °C, with continuous shaking at 900 r.p.m. with a diameter of 1 mm for K12, and with quiescent incubation with shaking 2 s before each reading for K19. Plots showing the fluorescence trace of the replicate with median lag time for each sample were created using R³⁹. Plots of lag time depict the mean time value at which each replicate crossed an arbitrary fluorescence value above noise background (values were selected per experiment and applied to all samples). Error bars represent the standard deviation of the replicate lag times for each sample.

Seeded K12 fibril formation assays. Seeds were produced by incubating 50 μ M K12 as above, but without ThS present, and were added at 0.25% (v/v). Peptide stock concentrations were 0.75 mM and were added at a final concentration of tenfold molar excess relative to soluble K12. Reaction mixtures were otherwise prepared and monitored as above.

VQIVYK inhibition assays. The VQIVYK fibril formation assay was modified from a previously published protocol⁴⁰. Buffers and plates were kept on ice to delay VQIVYK fibril formation while the reaction mixtures were prepared. Replicate solutions of 180 μ l of 25 mM MOPS, pH 7.2, 100 μ M ThS and inhibitor peptides were added to black, clear-bottomed, 96-well Nunc plates with 1/8-inch PTFE beads (Orange Products). Acetylated and amidated VQIVYK (Genscript) was dissolved in H₂O to 1.3 mM and filtered through a Millipore Microcon 100-kDa filter device at 14,000g for 5 min at 4 °C to remove large aggregates (final concentration, ~1 mM). Filtered VQIVYK (20 μ l) was added to each reaction well. ThS fluorescence was monitored at room temperature every 2 min using a SpectraMax M5 fluorometer with 2 s of mixing before each reading.

Amyloid- β fibril formation assay. Lyophilized amyloid- β (1–42) was diluted to 0.2 mg ml^{–1} in 50 mM NH₄OH and filtered with a 0.2- μ m filter. The reaction mixture contained a final concentration of 11.5 μ M amyloid- β (1–42), 10 μ M Thioflavin T (ThT), 23 mM NH₄OH in 100 mM bicine, pH 9.1, and 11.5 μ M D-TLKIVW in reactions with peptide present. Reactions were split into four replicates and the ThT fluorescence signal was measured every minute (excitation wavelength, 440 nm; emission wavelength, 510 nm), at 37 °C, with continuous shaking at 960 r.p.m. with a 1-mm diameter in a Varioskan fluorometer.

Electron microscopy. Sample (5 μ l) was applied to glow-discharged, 400-mesh carbon-coated, formvar films on copper grids (Ted Pella) for 3 min. Grids were rinsed twice with distilled water and stained with 1% uranyl acetate for 90 s. Grids were examined in a Hitachi H-7000 transmission electron microscope at 75 keV or a JEOL JEM1200-EX operating at 80 keV.

Tau fibril formation kinetic analysis. The nucleation (k_1) and propagation (k_2) rates were determined by fitting the form of the Finke–Watzky two-step mechanism⁴¹. Plateau values were determined and the remaining parameters were fitted using the 'leasqr' nonlinear least-squares regression function (<http://fly.isti.cnr.it/pub/software/octave/leasqr/>) through the OCTAVE software package (<http://www.gnu.org/software/octave/>).

Preparation of peptide–gold conjugates. Peptide–Nanogold conjugates were prepared as described earlier for similarly sized peptides⁴². Briefly, 60 nmol of the peptides CGGG-(D)-TLKIVW and CGGG-(D)-LKTWIV (CS Bio) were dissolved in 110 μ l of phosphate-buffered saline (20 mM, pH 6.5, 0.15 M NaCl), added to 6 nmol of Monomaleimido Nanogold (Nanoprobe), dissolved in 200 μ l H₂O and incubated for 1 h at room temperature (22 °C) with constant rotation. Peptide–Nanogold conjugates were separated from excess unbound peptides by membrane centrifugation (Microcon-10 system, Amicon) using a molecular mass cut-off of 10 kDa. Peptide–Nanogold conjugates were then diluted into phosphate-buffered saline, aliquoted and stored at –20 °C for no longer than one month.

Preparation of K19 fibrils. K19 fibrils were generated by incubating 100 μ M soluble K19 with 25 μ M 6-kDa heparin overnight at 37 °C in phosphate buffer (50 mM, pH 7.4). K19 fibrils were sonicated for 15 s, using a microtip set to 35% amplitude. Residual heparin and small oligomers were removed by centrifuging the mixture through a 100-kDa Microcon concentrator for 10 min at 14,000g, washing the retentate with phosphate buffer and repeating three times; the retentate was restored to its original volume with phosphate buffer. These short fibril segments were stored at 4 °C for no longer than one week. For NMR studies, fibril samples were similarly prepared, but were washed in H₂O and concentrated to 2 mM K19 (by monomer).

Preparation of samples for Nanogold binding experiments. Nanogold conjugated inhibitor (or control) (10 nM) was incubated with 1.67 μM K19 fibrils (by monomer) in MOPS buffer (25 mM, pH 7.2) for 1 h. We applied 5 μl of it to a glow-discharged, 400-mesh carbon-stabilized copper grid (Ted Pella) for 3 min. The grids were washed twice with H_2O and 10 μl of the Goldenhance reagent was applied for 10 s. The grids were washed five times with H_2O and negatively stained with 2% uranyl acetate.

Quantification and localization of Nanogold binding. For each sample, 75 Nanogold particles ≤ 15 nm in diameter were counted and classified as bound or unbound. The 15-nm cut-off was chosen to exclude unbound, but adjacent, particles enlarged by Goldenhance that only apparently bind fibrils. To establish the localization of the binding observed, individual Nanogold particles bound to fibrils were categorized as bound to the fibril end or side. In both of these experiments, sample identities were concealed from the microscopist to ensure unbiased counting. Grids were examined with a JEOL JEM1200-EX and images were recorded using DIGITALMICROGRAPH (Gatan).

Statistical analysis of Nanogold binding. We compared counts of Nanogold-conjugated peptides and unconjugated Nanogold bound to fibrils or localizing to fibril ends. Twenty-one unconjugated Nanogold particles out of 75 counted bound to fibrils. We modelled Nanogold particles bound to fibrils using a binomial distribution with parameters $n = 75$ (sample size: number of observations) and $P = 0.28$ (probability of success). In a separate experiment, 22 unconjugated Nanogold particles bound to fibrils that localized to fibril ends, following a binomial distribution with $n = 105$ and $P = 0.21$.

Because the number of counts is fairly large, we assumed a normal distribution and used a standard Z-test to compare the number of bound Nanogold-peptide conjugates with the expected distribution based on the number of bound, unconjugated Nanogold particles. We used an analogous analysis to determine the significance of localization to fibril ends.

The numbers of Nanogold-D-TLKIVW conjugates bound to fibrils ($x_{\text{bound}} = 43$, $n = 75$) and bound Nanogold-D-TLKIVW conjugates localizing to the end of fibrils ($x_{\text{end}} = 49$, $n = 86$) were significantly different from the corresponding numbers for Nanogold alone, whereas the number of Nanogold-D-LKTWIV conjugates bound ($x_{\text{bound}} = 15$, $n = 75$) or the number localized to fibril ends ($x_{\text{end}} = 17$, $n = 100$) did not differ significantly from the corresponding numbers for Nanogold alone.

VQIVYK preparation for binding studies. Acetylated and amidated VQIVYK peptide (Genscript) was dissolved to 1 mM in 25 mM MOPS, pH 7.2, and incubated at room temperature for at least 24 h. Fibrils were washed with H_2O , concentrated using an Amicon ultracentrifugal filter with a 3-kDa molecular mass cut-off and resuspended in H_2O to a final concentration (by monomer) of 4 mM. Soluble VQIVYK was prepared by dissolving VQIVYK peptide (CS Bio) with free amino and carboxy termini in H_2O .

^1H NMR sample preparation and measurements. NMR samples were prepared with 5% D_2O and 10 mM NaOAc, pH 5.0. D-peptides were added from 1 mM stocks in H_2O to a final concentration of 100 μM . Soluble and fibrillar VQIVYK and tau protein were added at indicated concentrations to make a final volume of 550 μl . ^1H NMR spectra measured at 500 MHz were collected on a Bruker DRX500 at 283 K. H_2O resonance was suppressed through presaturation. Spectra were processed with XWINNMR 3.6.

Binding constant estimations. NMR data were analysed to estimate a binding constant for the interaction between D-TLKIVW and VQIVYK fibrils. At about 1,000 μM VQIVYK (concentration as monomer), 50% of D-TLKIVW is bound (Supplementary Fig. 11). The steric-zipper model suggests that there are two monomers per 4.7 Å (0.47-nm) layer in a fibril¹⁴—such that the number of monomers per fibril is given by [fibril length (nm)] \times (2 monomers per 0.47 nm)—and we estimate the fibril concentration using the monomer concentration: $[\text{VQIVYK}_{\text{fibril}}] = [\text{VQIVYK}_{\text{monomer}}]/(\text{monomers per fibril})$. If we assume one binding site and estimate from electron microscopy an average length of ~ 140 nm per fibril, then there are about 600 monomers per fibril, and the apparent dissociation constant is about 2 μM .

Hydrogen-bonding energy calculation. We used AREAIMOL³¹ to calculate the non-polar and polar areas buried by the interaction between D-TLKIVW with the VQIVYK steric zipper (Fig. 2b, c and Supplementary Fig. 1). We calculate buried areas of 201, 24 and 102 Å² for carbon, nitrogen and oxygen atoms, respectively. Using the atomic solvation parameters of ref. 43, we estimate that the free energy of transferring the inhibitor from a non-polar phase to an aqueous phase, $\Delta G_{\text{solvation}}$, is approximately 2.5 kcal mol⁻¹. On the basis of an apparent dissociation constant of 2 μM , we estimate the total free energy change of bringing the inhibitor into contact with the VQIVYK steric-zipper template, $\Delta G_{\text{binding}}$, to be 7.4 kcal mol⁻¹. From the interaction model (Fig. 2c and Supplementary Fig. 1), we maintain six hydrogen bonds between D-TLKIVW and VQIVYK, and estimate the free energy change per hydrogen bond to be $(\Delta G_{\text{binding}} - \Delta G_{\text{solvation}})/6$, or ~ 0.8 kcal mol⁻¹.

GGVLVN crystallization and structure determination. The GGVLVN peptide was dissolved in 10 mM Tris, pH 9, at 1.8 mg ml⁻¹ and crystallized in 10% (w/v) PEG-8000, 0.1 M MES, pH 6.0, and 0.2 M Zn(OAc)₂. X-ray diffraction data was collected at APS beamline 24-ID-E. Phases were determined by molecular replacement using an idealized β -strand in PHASER⁴⁴. Crystallographic refinement was performed using REFMAC⁴⁵. Model building was performed with COOT⁴⁶ and illustrated with PYMOL⁴⁷.

$^{248}\text{PAP}^{286}$ fibril formation and inhibition. Fmoc- β -cyclohexyl-L-alanine and Fmoc-7-hydroxy-(S)-1,2,3,4-tetrahydroisoquinoline-3-carboxylic acid were purchased from AnaSpec and the inhibitor peptide Trp-His-Lys-chAla-Trp-hydroxyTic (WW61) was synthesized by Celtek Biosciences. $^{248}\text{PAP}^{286}$ and WW61 were dissolved as $\times 1.25$ and $\times 5$ stocks in PBS, respectively, and filtered with a 0.1- μm filter. $^{248}\text{PAP}^{286}$ was diluted with PBS to 0.66 mM and ThT was added to 10 μM final concentration. Samples were optionally mixed with 1.32 mM WW61 and vortexed. Five replicates of 150 μl were immediately dispensed into a 96-well plate. In dose-response experiments, WW61 final concentrations were 0.33, 0.66 and 1.32 mM. Plates were continuously agitated at 960 r.p.m. at 37 °C, and ThT fluorescence readings were recorded (excitation wavelength, 440 nm; emission wavelength, 482 nm) at 15-min intervals with a Varioskan Flash fluorometer. Lag time was determined when fluorescence crossed an arbitrary value (3 r.f.u.) above background.

Effect of WW61 on fibril-mediated enhancement of HIV-1 infection. The CCR5 tropic molecular HIV-1 clone NL4_3/92TH014-2⁴⁸ was generated by transient transfection of 293T cells with proviral DNA. Supernatants were collected 48 h later and p24 concentrations determined by ELISA. TZM-bl reporter cells encoding a *lacZ* gene under the control of the viral LTR promoter were obtained through the NIH AIDS Research and Reference Reagent Program and provided by Dr John C. Kappes, Dr Xiaoyun Wu and Tranzyme⁴⁹. HIV-1 (40 μl) containing 0.1 ng of p24 antigen was incubated with 40- μl dilutions of mixtures of $^{248}\text{PAP}^{286}$ and inhibitory peptide, WW61, that was either freshly prepared or had been agitated for 23 h. Peptide concentrations and experimental conditions during agitation were similar to those described above. Thereafter, 20 μl of the mixtures were used to infect 180 μl of TZM-bl cells seeded the day before (10^5 per well). Two days later, infection rates were determined by quantifying β -galactosidase activities in cellular lysates using the Gal-Screen assay (Applied Biosystems, T1027). Luminescence was recorded on an Orion microplate luminometer as relative light units per second.

Effect of WW61 on polylysine-mediated enhancement of HIV-1 infection. Polylysine (Sigma Aldrich) (50 μl) was mixed with an equal volume of WW61. Thereafter, 35- μl fivefold dilutions of the polylysine-WW61 mixture or polylysine alone were incubated with the same volume of virus and incubated for 5 min at room temperature. Polylysine-WW61 concentrations were 100, 20, 4, 0.8, 0.16, 0.032, 0.064 and 0 $\mu\text{g ml}^{-1}$ during pre-incubation with virus stocks. Thereafter, 20 μl of each mixture was added to 180 μl of TZM-bl cells. The infection rate was determined two days later as described above.

Effect of WW61, GIHKQK and PYKLWN on HIV-1 infection. Each peptide (40 μl) was incubated with an equal volume of virus containing 1 ng of p24 antigens for 5 min at room temperature. Peptide concentrations were 150, 30, 6, 1.2 and 0 $\mu\text{g ml}^{-1}$ during pre-incubation with virus stocks. Thereafter, 20 μl of each mixture was added separately to 180 μl of TZM-bl cells (tenfold dilution) and the infection rate was determined as above.

- Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
- Lawrence, M. C. & Colman, P. M. Shape complementarity at protein/protein interfaces. *J. Mol. Biol.* **234**, 946–950 (1993).
- Studier, F. W., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. W. Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol.* **185**, 60–89 (1990).
- Biernat, J. *et al.* The switch of tau protein to an Alzheimer-like state includes the phosphorylation of two serine-proline motifs upstream of the microtubule binding region. *EMBO J.* **11**, 1593–1597 (1992).
- Barghorn, S., Biernat, J. & Mandelkow, E. Purification of recombinant tau protein and preparation of Alzheimer-paired helical filaments in vitro. *Methods Mol. Biol.* **299**, 35–51 (2005).
- Friedhoff, P., Schneider, A., Mandelkow, E. M. & Mandelkow, E. Rapid assembly of Alzheimer-like paired helical filaments from microtubule-associated protein tau monitored by fluorescence in solution. *Biochemistry* **37**, 10223–10230 (1998).
- Pérez, M., Valpuesta, J. M., Medina, M., Montejó de Garcini, E. & Avila, J. Polymerization of tau into filaments in the presence of heparin: the minimal sequence required for tau-tau interaction. *J. Neurochem.* **67**, 1183–1190 (1996).
- Schweers, O., Mandelkow, E. M., Biernat, J. & Mandelkow, E. Oxidation of cysteine-322 in the repeat domain of microtubule-associated protein tau controls the in vitro assembly of paired helical filaments. *Proc. Natl Acad. Sci. USA* **92**, 8463–8467 (1995).

39. R Development Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna) (<http://www.r-project.org>) (2008).
40. Rojas Quijano, F. A., Morrow, D., Wise, B. M., Brancia, F. L. & Goux, W. J. Prediction of nucleating sequences from amyloidogenic propensities of tau-related peptides. *Biochemistry* **45**, 4638–4652 (2006).
41. Morris, A. M., Watzky, M. A., Agar, J. N. & Finke, R. G. Fitting neurological protein aggregation kinetic data via a 2-step, minimal “Ockham’s razor” model: the Finke-Watzky mechanism of nucleation followed by autocatalytic surface growth. *Biochemistry* **47**, 2413–2427 (2008).
42. Schmidt, K., Segond von Banchet, G. & Heppelmann, B. Labelling of peptides with 1.4-nm gold particles to demonstrate their binding sites in the rat spinal cord. *J. Neurosci. Methods* **87**, 195–200 (1999).
43. Eisenberg, D., Wesson, M. & Yamashita, M. Interpretation of protein folding and binding with atomic solvation parameters. *Chem. Scr.* **29A**, 217–221 (1989).
44. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
45. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).
46. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
47. DeLano, W. L. *PyMOL Molecular Viewer* (<http://www.pymol.org>) (2002).
48. Papkalla, A., Munch, J., Otto, C. & Kirchhoff, F. Nef enhances human immunodeficiency virus type 1 infectivity and replication independently of viral coreceptor tropism. *J. Virol.* **76**, 8455–8459 (2002).
49. Platt, E. J., Wehrly, K., Kuhmann, S. E., Chesebro, B. & Kabat, D. Effects of CCR5 and CD4 cell surface concentrations on infections by macrophagetropic isolates of human immunodeficiency virus type 1. *J. Virol.* **72**, 2855–2864 (1998).

Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia

Xose S. Puente¹, Magda Pinyol², Víctor Quesada¹, Laura Conde³, Gonzalo R. Ordóñez¹, Neus Villamor³, Georgia Escaramis⁴, Pedro Jares³, Silvia Beà³, Marcos González-Díaz⁵, Laia Bassaganyas⁴, Tycho Baumann⁶, Manel Juan⁷, Mónica López-Guerra³, Dolors Colomer³, José M. C. Tubío^{4,8}, Cristina López³, Alba Navarro³, Cristian Tornador⁴, Marta Aymerich³, María Rozman³, Jesús M. Hernández⁵, Diana A. Puente¹, José M. P. Freije¹, Gloria Velasco¹, Ana Gutiérrez-Fernández¹, Dolors Costa³, Anna Carrió³, Sara Guijarro³, Anna Enjuanes³, Lluís Hernández³, Jordi Yagüe⁷, Pilar Nicolás⁹, Carlos M. Romeo-Casabona⁹, Heinz Himmelbauer¹⁰, Ester Castillo¹⁰, Juliane C. Dohm¹⁰, Silvia de Sanjosé¹¹, Miguel A. Piris¹², Enrique de Alava⁵, Jesús San Miguel⁵, Romina Royo¹³, Josep L. Gelpi¹³, David Torrents¹³, Modesto Orozco¹³, David G. Pisano¹⁴, Alfonso Valencia¹⁴, Roderic Guigó¹⁵, Mónica Bayés¹⁶, Simon Heath¹⁶, Marta Gut¹⁶, Peter Klatt¹⁷, John Marshall¹⁸, Keiran Raine¹⁸, Lucy A. Stebbings¹⁸, P. Andrew Futreal¹⁸, Michael R. Stratton¹⁸, Peter J. Campbell¹⁸, Ivo Gut¹⁶, Armando López-Guillermo⁶, Xavier Estivill⁴, Emili Montserrat⁶, Carlos López-Otín^{1*} & Elías Campo^{3*}

Chronic lymphocytic leukaemia (CLL), the most frequent leukaemia in adults in Western countries, is a heterogeneous disease with variable clinical presentation and evolution^{1,2}. Two major molecular subtypes can be distinguished, characterized respectively by a high or low number of somatic hypermutations in the variable region of immunoglobulin genes^{3,4}. The molecular changes leading to the pathogenesis of the disease are still poorly understood. Here we performed whole-genome sequencing of four cases of CLL and identified 46 somatic mutations that potentially affect gene function. Further analysis of these mutations in 363 patients with CLL identified four genes that are recurrently mutated: notch 1 (*NOTCH1*), exportin 1 (*XPO1*), myeloid differentiation primary response gene 88 (*MYD88*) and kelch-like 6 (*KLHL6*). Mutations in *MYD88* and *KLHL6* are predominant in cases of CLL with mutated immunoglobulin genes, whereas *NOTCH1* and *XPO1* mutations are mainly detected in patients with unmutated immunoglobulins. The patterns of somatic mutation, supported by functional and clinical analyses, strongly indicate that the recurrent *NOTCH1*, *MYD88* and *XPO1* mutations are oncogenic changes that contribute to the clinical evolution of the disease. To our knowledge, this is the first comprehensive analysis of CLL combining whole-genome sequencing with clinical characteristics and clinical outcomes. It highlights the usefulness of this approach for the identification of clinically relevant mutations in cancer.

To gain insights into the molecular alterations that cause CLL, we performed whole-genome sequencing of four cases representative of different forms of the disease: two cases, CLL1 and CLL2, with no mutations in the immunoglobulin genes (*IGHV*-unmutated) and two cases, CLL3 and CLL4, with mutations in these genes (*IGHV*-mutated) (Supplementary Table 1 and Supplementary Information). We used a combination of whole-genome sequencing and exome sequencing, as well as long-insert paired-end libraries, to detect variants in chromosomal structure (Supplementary Fig. 1 and Supplementary Tables 2–5).

We obtained more than 99.7% concordance between whole-genome sequencing calls and genotyping data, indicating that the coverage and parameters used were sufficient to detect most of the sequence variants in these samples (Supplementary Information). We detected about 1,000 somatic mutations per tumour in non-repetitive regions (Fig. 1a, Supplementary Fig. 2 and Supplementary Table 6). These numbers of somatic mutations were lower than the numbers in melanoma and lung carcinoma^{5,6}, but in agreement with previous estimates of less than one mutation per megabase (Mb) for leukaemias⁷. The most common substitution was the transition G>A/C>T, usually occurring in a CpG context (Fig. 1b and Supplementary Fig. 2). We also detected marked differences in the mutation pattern between CLL samples and these differences were associated with tumour subtype (Fig. 1b). Thus, *IGHV*-mutated cases showed a higher proportion of A>C/T>G mutations than cases with unmutated *IGHV* ($16 \pm 0.2\%$ versus $6.2 \pm 0.1\%$). The base preceding the adenine in A to C transversions showed an overrepresentation of thymine, when compared to the prevalence expected from its representation in non-repetitive sequences in the wild-type genome ($P < 0.001$, Fig. 1c), and there were fewer A to C substitutions at GpA dinucleotides than would be expected by chance ($P < 0.001$). These differences between CLL subtypes might reflect the molecular mechanisms implicated in their respective development. The pattern and context of mutations are consistent with their being introduced by the error-prone polymerase η during somatic hypermutation in immunoglobulin genes⁸. This indicates that polymerase η could contribute to the high frequency of A > T to C > G transversions in cases with *IGHV*-mutated. It also extends the differences observed between these two CLL subtypes to the genomic level.

We classified the somatic mutations into three different classes according to their potential functional effect (Supplementary Information). We also searched for small insertions and deletions (indels) in coding regions: we found and validated five somatic indels, which caused frameshifts in protein-coding regions (Supplementary Table 7).

¹Departamento de Bioquímica y Biología Molecular, Instituto Universitario de Oncología, Universidad de Oviedo, 33006 Oviedo, Spain. ²Unidad de Genómica, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), 08036 Barcelona, Spain. ³Unidad de Hematopatología, Servicio de Anatomía Patológica, Hospital Clínic, Universitat de Barcelona, IDIBAPS, 08036 Barcelona, Spain. ⁴Genes and Disease Programme, Center for Genomic Regulation, Pompeu Fabra University (CRG-UPF), 08003 Barcelona, Spain. ⁵Servicio de Hematología, Hospital Universitario, Centro de Investigación del Cáncer-IBMC (USAL-CSIC), Universidad de Salamanca, 37007 Salamanca, Spain. ⁶Servicio de Hematología, Hospital Clínic, IDIBAPS, Universidad de Barcelona, 08036 Barcelona, Spain. ⁷Servicio de Inmunología, Hospital Clínic, IDIBAPS, 08036 Barcelona, Spain. ⁸Hospital Clínico Universitario de Santiago de Compostela, 15706 Santiago de Compostela, Spain. ⁹Cátedra Inter-Universitaria de Derecho y Genoma Humano, Universidad de Deusto, Universidad del País Vasco, 48007 Bilbao, Spain. ¹⁰UltraSequencing Unit, Center for Genomic Regulation, Pompeu Fabra University (CRG-UPF), 08003 Barcelona, Spain. ¹¹Unidad de Infecciones y Cáncer, Institut Català de Oncologia-IDIBELL, 08907 Hospitalet de Llobregat, Spain. ¹²Molecular Pathology Programme, Spanish National Cancer Research Centre (CNIO), 28029 Madrid, Spain. ¹³Programa Conjunto de Biología Computacional, Barcelona Supercomputing Center (BSC), Institut de Recerca Biomèdica (IRB), Spanish National Bioinformatics Institute, Universitat de Barcelona, 08028 Barcelona, Spain. ¹⁴Structural Biology and Biocomputing Programme, Spanish National Cancer Research Centre (CNIO), Spanish National Bioinformatics Institute, 28029 Madrid, Spain. ¹⁵Genomic Bioinformatics Programme, Center for Genomic Regulation, Pompeu Fabra University (CRG-UPF), Spanish National Bioinformatics Institute, 08003 Barcelona, Spain. ¹⁶Centro Nacional de Análisis Genómico, Parc Científic de Barcelona, 08028 Barcelona, Spain. ¹⁷Centro Nacional de Biotecnología, Consejo Superior de Investigaciones Científicas, 28049 Madrid, Spain. ¹⁸Wellcome Trust Sanger Institute, Hinxton CB10 1SA, UK.

*These authors contributed equally to this work.

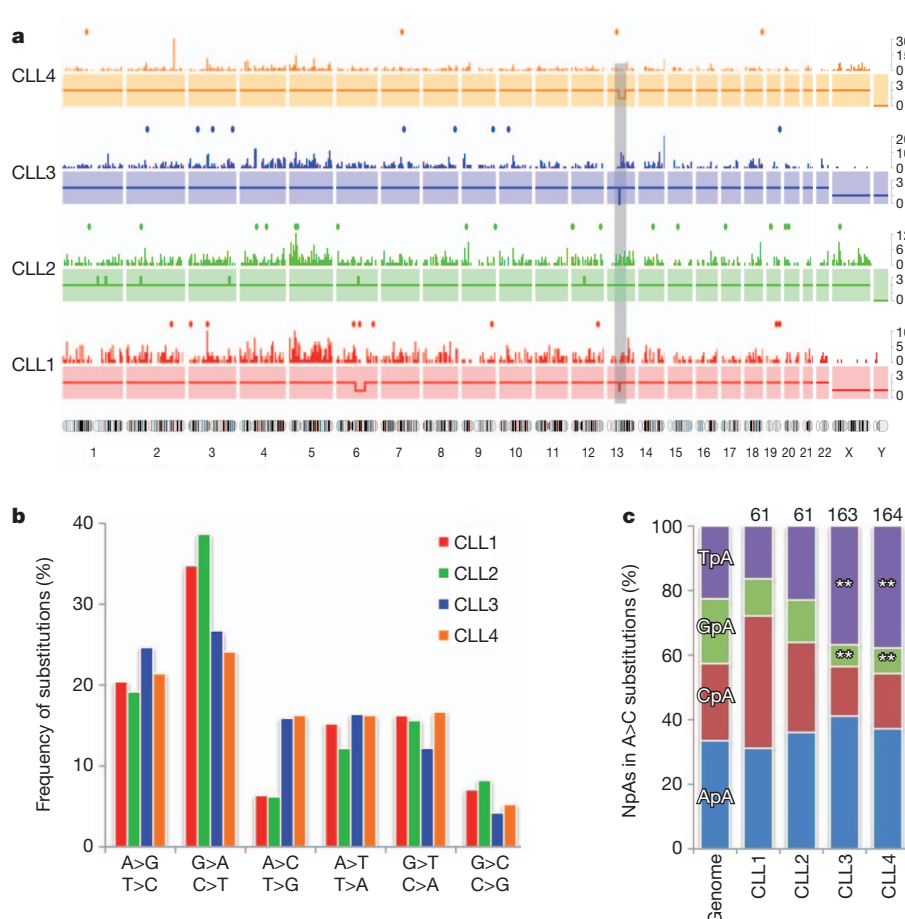


Figure 1 | Profile of somatic mutations in four CLL genomes. **a**, Distribution of somatic alterations. For each tumour genome, copy number (solid lines), density of mutations per 5-Mb window (bars) and protein-coding mutations (dots) are shown. The shaded rectangle indicates the location of the 13q14 deletion that was present in three of the four CLL cases. Chromosome numbers

are listed below the four profiles. **b**, Frequency of substitutions in each CLL tumour for the six possible classes of mutation. **c**, Distribution of the four possible NpA dinucleotides for the A to C transversion in each tumour genome, compared with the expected distribution across the genome. The total number of A to C substitutions per case is indicated at the top (**, $P < 0.001$).

We identified 46 mutations that changed the protein-coding sequences of 45 genes in the four patients analysed (Supplementary Table 7). None of these nucleotide substitutions had been previously linked to CLL and among the five indel mutations, only one, in *NOTCH1* (p.P2515Rfs*4), had been previously found in various lymphoid malignancies, including CLL^{9,10}. To determine whether any of these 45 genes was mutated in more than one CLL case, we analysed an initial validation set of 169 CLL patients. We focused on the 26 genes that are expressed at the RNA level in CLL cells (Supplementary Table 7) because mutations in expressed genes are more likely to have a biological effect than those in non-expressed genes. We used a pooled-sequencing strategy that led us to identify four genes with at least one additional mutation in the validation series: these were *NOTCH1*, *MYD88*, *XPO1* and *KLHL6* (Table 1 and Supplementary Information).

Analysis of additional CLL cases revealed that the deletion of a CT dinucleotide in *NOTCH1* (p.P2515Rfs*4) was found in 29 of 255 patients and two additional mutations in the same region were also found (p.Q2503* and p.F2482Ffs*2) (Fig. 2a, b). Accordingly, *NOTCH1* is mutated in 12% of CLL patients (Supplementary Table 8). These mutations generate a premature stop codon, resulting in a *NOTCH1* protein lacking the C-terminal domain, which contains a PEST sequence (a sequence rich in proline, glutamic acid, serine and threonine) (Fig. 2a). Removal of this region results in the accumulation of an active protein isoform in the mutated CLL cells (Fig. 2c and Supplementary Fig. 3). *NOTCH1* is constitutively expressed in CLL¹¹, but the *NOTCH1* mutations identified herein generate a more stable and active isoform of the protein. Gene expression analysis of ten *NOTCH1*-mutated and 49 unmutated CLL cases revealed a high

Table 1 | Genes recurrently mutated in chronic lymphocytic leukaemia

Gene	Protein	Mutation	Mutated cases / total	Overall frequency (%)	Frequency in IGHV-unmutated (%)	Frequency in IGHV-mutated (%)
<i>NOTCH1</i>	Notch 1	P2515Rfs*4 Q2503* F2482Ffs*2	29/255 1/255 1/255	12.2	20.4	7
<i>MYD88</i>	Myeloid differentiation primary response gene 88	L265P	9/310	2.9	0.8	5.6
<i>XPO1</i>	Exportin 1	E571K E571G	3/165 1/165	2.4	4.6	0
<i>KLHL6</i>	Kelch-like 6	F49L/L65P L90F L58P/T64A/Q81P	3/160	1.8	0	4.5

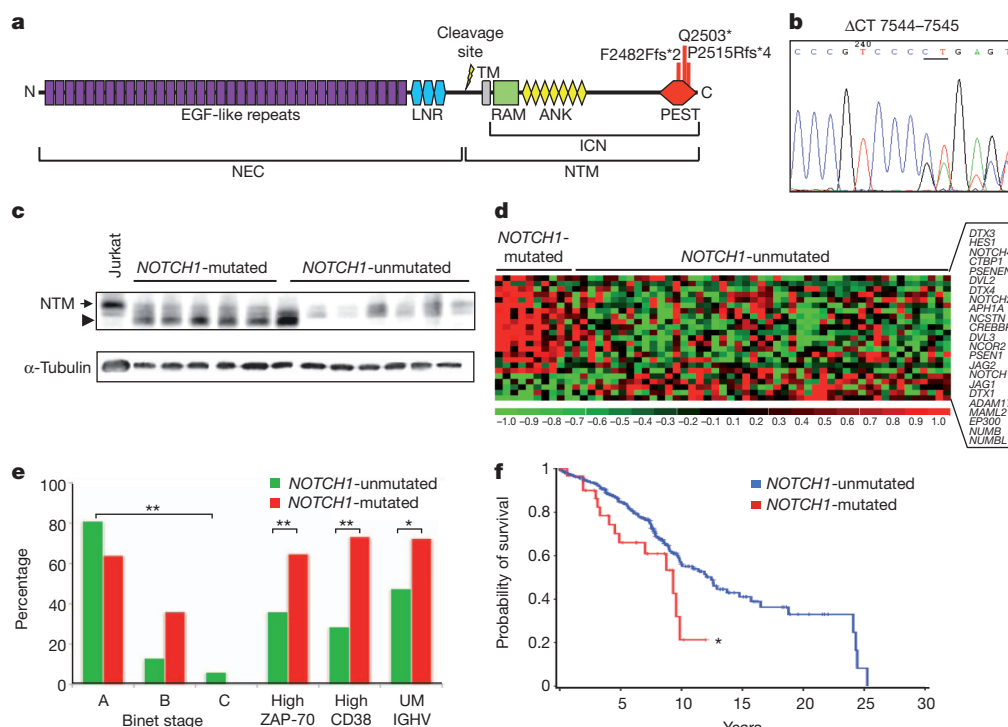


Figure 2 | Mutational and functional analysis of NOTCH1 in CLL.

a, Schematic representation of human NOTCH1, showing the main domains and locations of the three different somatic mutations identified in CLL. NEC, NOTCH1 extracellular subunit; NTM, NOTCH1 transmembrane subunit; ICN, intracellular domain of NOTCH1; LNR, Lin-12 NOTCH repeats; RAM, RAM domain; ANK, ankyrin repeat domain; PEST, PEST domain. **b**, Electropherogram showing the heterozygous CT deletion recurrently identified in CLL. **c**, Western blot showing NOTCH1 protein levels in CLL cases with or without the NOTCH1 p.P2515Rfs*4 mutation, and in Jurkat cells

as a control. The arrow indicates the band corresponding to the NTM; the large arrowhead indicates the smaller band corresponding to the mutant form.

d, Heat map showing the 23 genes of the NOTCH1 pathway that are differentially expressed in NOTCH1-mutated versus non-mutated CLL.

e, Distribution of disease stage (Binet), ZAP-70 expression status, CD38 expression status and IGHV mutational status (UM, unmutated IGHV) in patients with or without mutations in NOTCH1 (*, $P < 0.02$; **, $P < 0.01$).

f, Actuarial probability of overall survival of CLL patients with mutated or unmutated NOTCH1 (*, $P = 0.03$).

number of differentially expressed genes ($n = 542$, false discovery rate < 0.05 ; Supplementary Table 9). Likewise, in a gene-set analysis, we found that there was significant differential expression of the NOTCH1 signalling pathway¹² and two metabolic pathways (oxidative phosphorylation and glycolysis/gluconeogenesis). This is consistent with the NOTCH1-mediated activation of multiple biosynthetic routes in T acute lymphoblastic leukaemia¹³. When the differential expression of individual genes from the NOTCH1 pathway was analysed, 23 of the 46 genes assigned to this pathway¹² showed a significant differential expression ($P < 0.05$) in NOTCH1-mutated CLL (Fig. 2d). NOTCH1-mutated patients had a more advanced clinical stage at diagnosis, more adverse biological features and an overall survival that was significantly shorter than those with NOTCH1 unmutated (10-yr overall survival: 21% versus 56%, $P = 0.03$; Fig. 2e, f). NOTCH1-mutated CLL also underwent transformation into diffuse large B-cell lymphoma more frequently than NOTCH1-unmutated CLL (7 of 31 cases, 23%, versus 3 of 224 cases, 1.3%; $P < 0.001$). The same IGHV clonal rearrangement and NOTCH1 mutation were found in the CLL and corresponding transformed diffuse large B-cell lymphoma of the four cases studied, indicating a clonal relationship of both components.

A recurrent mutation (p.L265P) in the MYD88 gene (Fig. 3a, b) was also identified in 9 of 310 CLL patients (2.9%). During revision of this manuscript, the same mutation has been identified in different lymphomas¹⁴, highlighting its relevance in the pathogenesis of lymphoid neoplasias. This protein participates in the signalling pathways of interleukin-1 and Toll-like receptors during the immune response¹⁵. MyD88 immunoprecipitation from CLL cells with the p.L265P mutation resulted in the co-immunoprecipitation of large amounts of IRAK1, in contrast to cells lacking this mutation (Fig. 3c). Other effectors of this signalling pathway, including STAT3, IκBα and NF-κB p65

subunit, showed higher phosphorylation in MYD88-mutated than in unmutated CLL cells (Fig. 3d, e) and there was an increased DNA-binding activity of NF-κB in MYD88-mutated cells (Supplementary Fig. 4). These data support the hypothesis that the MYD88 p.L265P mutation constitutes an activating mutation of this novel proto-oncogene^{14,16}. Stimulation of interleukin-1 receptor or Toll-like receptors in MYD88-mutated CLL cells induced the secretion of 5-fold to 150-fold higher levels of interleukin 1 receptor antagonist (IL1RN, also known as IL1RA), interleukin 6 and chemokine (C-C motif) ligands 2, 3 and 4 (CCL2, CCL3 and CCL4), when compared to the secretion of these cytokines by MYD88-unmutated CLLs. Cytokine secretion was elevated in MYD88-mutated cells in response to stimulation of at least four of the eight TLRs tested. No response was observed in lymphocytes carrying the inactivating MYD88 mutation E52DEL (Fig. 3f and Supplementary Fig. 5). The high production of these cytokines has been implicated in the recruitment of macrophages and T lymphocytes by CLL cells, creating a favourable niche for their survival¹⁷. Moreover, activation of Toll-like receptors in CLL cells promotes the proliferation of tumour cells and protects them from spontaneous apoptosis¹⁸. Patients with MYD88-mutated CLL were diagnosed at a younger age than those with wild-type MYD88 (median 43 yr, range 38–63, versus median 63 yr, range 27–94; $P < 0.001$) and the disease presented with a more advanced clinical stage (Fig. 3g), although no differences were observed in progression or survival rates. Notably, almost all patients with the MYD88 p.L265P mutation (seven of the eight evaluated) belonged to the IGHV-mutated group.

We also identified four cases with mutations in the same codon of the exportin 1 gene (XPO1; p.E571K and p.E571G). Exportin 1 is implicated in the nuclear export of proteins and mRNAs in yeast, including members of the MAP kinase pathway¹⁹. The fact that the

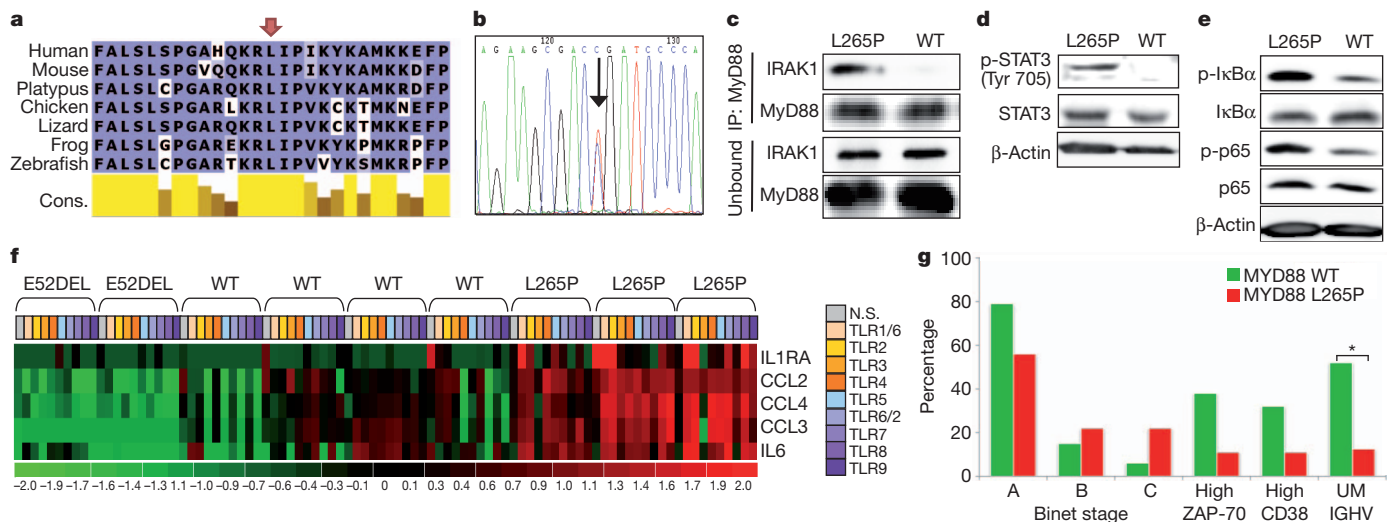


Figure 3 | Mutational and functional analysis of MYD88 in CLL. **a**, Multiple sequence alignment of MYD88 around the mutated residue (arrow) in different species. Cons., degree of conservation. **b**, Electropherogram showing the recurrent heterozygous p.L265P MYD88 mutation (arrow) detected in CLL. **c**, Cell extracts from a MYD88-mutated CLL (L265P) and a MYD88-unmutated CLL (WT) were immunoprecipitated with anti-MyD88 antibody. The immunoprecipitated and unbound fractions were analysed by western blot using anti-IRAK1 and anti-MyD88 antibodies. **d**, Western blot analysis of phosphorylated STAT3 (p-STAT3 (Tyr 705)) and total STAT3 in cell extracts from MYD88-mutated or unmutated CLL tumour cells. β-Actin was used as a control to show equal loading. **e**, Western blots showing phosphorylated IκBα (p-IκBα), total IκBα, phosphorylated p65 subunit of NF-κB (p-p65) and total p65 subunit of NF-κB in cell extracts from MYD88-mutated or unmutated CLL

same residue is mutated in four CLL cases and is part of a highly conserved region (Supplementary Fig. 6) indicates that the mutation affects XPO1 activity. Notably, all four cases with mutations in *XPO1* belonged to the *IGHV*-unmutated subtype and two of them also had the p.P2515Rfs*4 mutation in *NOTCH1*, indicating that both mutations could have synergic effects in CLL development.

We identified three patients carrying a total of six mutations (F49L/L65P, L90F and L58P/T64A/Q81P) in the gene encoding kelch-like protein 6 (*KLHL6*), which is implicated in the formation of the germinal centre during B cell maturation²⁰. All six mutations were clustered between residues 49 and 90 (Supplementary Fig. 7). The presence of several point mutations in *cis*, located near the transcriptional start site of a gene that is highly expressed in the germinal centre, is a characteristic feature of somatic hypermutation. In fact, all three patients had CLL with mutated *IGHV*. Although somatic hypermutation occurs mainly in *IGHV* regions, other proto-oncogenes, including *BCL6*, *MYC* and *PIM1*, are mutated by somatic hypermutation in different lymphomas²¹. However, only *BCL6* has been previously shown to be hypermutated by this mechanism in CLL²¹. Our data show that *KLHL6* is probably also a target of somatic hypermutation in *IGHV*-mutated patients, although its precise contribution to the oncogenic process in CLL remains to be determined.

In addition to these four genes, we identified a series of large genomic alterations that were previously reported². They included the deletion, in three cases, of the 13q14 region²², and a 40-Mb deletion in chromosome 6q14–q22 (Fig. 1a, Supplementary Fig. 1 and Supplementary Table 5). Finally, in one patient we detected a p.P281R mutation in the cyclin D2 gene (*CCND2*), which resulted in the accumulation of cyclin D2 in tumour cells (Supplementary Fig. 8). This finding, together with the high conservation of this residue and the identification of mutations in the equivalent residue of cyclin D1 (*CCND1*) in endometrial cancer²³, indicates that this *CCND2* mutation could be a driver contributing to the development of CLL in this patient. The finding illustrates the putative relevance of non-recurrent mutations for the pathogenesis of CLL.

tumour cells. **f**, Heat map representing the cytokine levels secreted by B cells from eight individuals after Toll-like receptor stimulation. Only the cytokines that showed the most significant differences between MYD88-mutated and MYD88-unmutated CLL are shown. 'E52DEL' indicates B cells from two patients with an inactivating MYD88 mutation, 'WT' corresponds to tumour cells from CLL patients without MYD88 mutation and 'L265P' indicates tumour cells from patients carrying a mutated MYD88. The stimulation experiments for each of the Toll-like receptors (TLRs) are represented in different colours. NS, no stimulus. **g**, Distribution of disease stage (Binet), ZAP-70 expression status, CD38 expression status and *IGHV* mutational status (UM, unmutated *IGHV*) in patients according to the presence or absence of p.L265P MYD88 mutation (*, $P < 0.03$).

The International Cancer Genome Consortium project was founded on the concept that sequencing of cancer genomes could reshape our understanding of cancer biology, with direct implications for clinical translation²⁴. Our study of four CLL genomes underscores this transformative potential, although additional studies will be necessary to translate these findings to the clinic. We have identified four recurrently mutated genes and provided novel insights into the mechanisms by which leukaemic cells recruit, instruct and coordinate a tumour microenvironment. Currently, the biological identification of different subgroups of CLL is based on markers such as *IGHV* mutational status, cytogenetics, ZAP-70 expression or CD38 expression, which are not fundamental agents in the leukaemic process. The classification of patients based on genomic drivers of the disease is conceptually appealing, as shown by our demonstration that *NOTCH1* and *MYD88* mutations identify distinct subgroups of patients with particular clinical and biological features. Furthermore, we provide functional evidence that both *NOTCH1* and *MYD88* mutations are activating events and potential therapeutic targets. The potential to personalize therapeutic choices for patients on the basis of the genomic architecture of their cancers is the long-term aspiration for studies such as this, combining whole-genome sequencing, functional studies and clinical analysis of patients with cancer.

METHODS SUMMARY

Four patients with CLL, who had given informed consent for sample collection and analysis, were studied. Tumour samples were obtained before treatment and tumour cells were separated from non-tumour cells by immunomagnetic depletion of T cells, natural killer cells, monocytes and granulocytes (Supplementary Information). Tumour cell purity was $\geq 98\%$ as assessed by flow cytometry. Normal blood cells from the same patient were obtained after treatment, resulting in no detectable, or less than 0.05%, tumour cell contamination, as assessed by flow cytometry. Additional samples from 363 patients were obtained for clinical validation. Protocols for long-insert and short-insert library construction and for massively parallel paired-end sequencing have been described elsewhere (ref. 25 and Supplementary Information). Genotyping and copy number analysis were performed using the Affymetrix SNP6.0, Agilent 1M and Illumina OmniQuad arrays

on the same cases used for whole-genome sequencing. For the validation of candidate genes in a set of 169 additional CLL patients, we used a combination of PCR amplification and Illumina sequencing in pooled samples, resulting in efficient identification of germline and somatic mutations (Supplementary Information). Sequencing data were aligned to the human reference genome (GRCh37) using Burrows–Wheeler alignment (BWA)²⁶ and somatic substitutions were identified using Sidrón, a probabilistic binomial model that uses genotyping data to calibrate sequencing error per sample. Functional analyses of the identified mutations were performed using cryopreserved primary tumour cells. For gene expression analysis, RNA was purified from tumour cells and analysed using the HU133 plus 2.0 GeneChip (Affymetrix). For immunoprecipitation and western blotting, CLL cell extracts were prepared and detected using the indicated antibodies (Supplementary Information). For Toll-like receptor stimulation of CLL cells, the Human TLR1–9 agonist kit (InvivoGen) was used.

Received 16 November 2010; accepted 6 April 2011.

Published online 5 June 2011.

- Rozman, C. & Montserrat, E. Chronic lymphocytic leukemia. *N. Engl. J. Med.* **333**, 1052–1057 (1995).
- Zenz, T., Mertens, D., Kuppers, R., Dohner, H. & Stilgenbauer, S. From pathogenesis to treatment of chronic lymphocytic leukaemia. *Nature Rev. Cancer* **10**, 37–50 (2010).
- Damle, R. N. *et al.* Ig V gene mutation status and CD38 expression as novel prognostic indicators in chronic lymphocytic leukemia. *Blood* **94**, 1840–1847 (1999).
- Hamblin, T. J., Davis, Z., Gardiner, A., Oscier, D. G. & Stevenson, F. K. Unmutated Ig V(H) genes are associated with a more aggressive form of chronic lymphocytic leukemia. *Blood* **94**, 1848–1854 (1999).
- Pleasant, E. D. *et al.* A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* **463**, 191–196 (2010).
- Pleasant, E. D. *et al.* A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature* **463**, 184–190 (2010).
- Greenman, C. *et al.* Patterns of somatic mutation in human cancer genomes. *Nature* **446**, 153–158 (2007).
- Spencer, J. & Dunn-Walters, D. K. Hypermutation at A-T base pairs: the A nucleotide replacement spectrum is affected by adjacent nucleotides and there is no reverse complementarity of sequences flanking mutated A and T nucleotides. *J. Immunol.* **175**, 5170–5177 (2005).
- Weng, A. P. *et al.* Activating mutations of *NOTCH1* in human T cell acute lymphoblastic leukemia. *Science* **306**, 269–271 (2004).
- Sportoletti, P. *et al.* *NOTCH1* PEST domain mutation is an adverse prognostic factor in B-CLL. *Br. J. Haematol.* **151**, 404–406 (2010).
- Rosati, E. *et al.* Constitutively activated Notch signaling is involved in survival and apoptosis resistance of B-CLL cells. *Blood* **113**, 856–865 (2009).
- Kanehisa, M., Goto, S., Furumichi, M., Tanabe, M. & Hirakawa, M. KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res.* **38**, D355–D360 (2010).
- Palomero, T. *et al.* *NOTCH1* directly regulates *c-MYC* and activates a feed-forward-loop transcriptional network promoting leukemic cell growth. *Proc. Natl Acad. Sci. USA* **103**, 18261–18266 (2006).
- Ngo, V. N. *et al.* Oncogenically active *MYD88* mutations in human lymphoma. *Nature* **470**, 115–119 (2011).
- O'Neill, L. A. & Bowie, A. G. The family of five: TIR-domain-containing adaptors in Toll-like receptor signalling. *Nature Rev. Immunol.* **7**, 353–364 (2007).
- Coste, I. *et al.* Dual function of MyD88 in RAS signaling and inflammation, leading to mouse and human cell transformation. *J. Clin. Invest.* **120**, 3663–3667 (2010).
- Burger, J. A. *et al.* High-level expression of the T-cell chemokines CCL3 and CCL4 by chronic lymphocytic leukemia B cells in nurse-like cell cocultures and after BCR stimulation. *Blood* **113**, 3050–3058 (2009).
- Muzio, M. *et al.* Expression and function of toll like receptors in chronic lymphocytic leukaemia cells. *Br. J. Haematol.* **144**, 507–516 (2009).
- Ferrigno, P., Posas, F., Koepp, D., Saito, H. & Silver, P. A. Regulated nucleo/cytoplasmic exchange of HOG1 MAPK requires the importin β homologs NMD5 and XPO1. *EMBO J.* **17**, 5606–5614 (1998).
- Kroll, J. *et al.* The BTB-kelch protein KLHL6 is involved in B-lymphocyte antigen receptor signaling and germinal center formation. *Mol. Cell. Biol.* **25**, 8531–8540 (2005).
- Pasqualucci, L. *et al.* Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. *Nature* **412**, 341–346 (2001).
- Aqeilan, R. I., Calin, G. A. & Croce, C. M. miR-15a and miR-16-1 in cancer: discovery, function and future perspectives. *Cell Death Differ.* **17**, 215–220 (2010).
- Moreno-Bueno, G. *et al.* Cyclin D1 gene (*CCND1*) mutations in endometrial cancer. *Oncogene* **22**, 6115–6118 (2003).
- Hudson, T. J. *et al.* International network of cancer genome projects. *Nature* **464**, 993–998 (2010).
- Bentley, D. R. *et al.* Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**, 53–59 (2008).
- Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was funded by the Spanish Ministry of Science and Innovation (MICINN) through the Instituto de Salud Carlos III (ISCIII) and Red Temática de Investigación del Cáncer (RTICC) del ISCIII. C.L.-O. is an Investigator of the Botín Foundation and D.T., of the ICREA program. We thank E. Santos for his support of this project, A. Carracedo and J. Benítez for genotyping studies, C. Fortuny for the supply of samples and N. Villahoz and M. C. Muro for their work in the coordination of the CLL-ICGC Consortium. We are also grateful to all patients with CLL who participated in this study.

Author Contributions X.S.P., V.Q., G.R.O., J.M.P.F., G.V., and A.G.-F. developed the bioinformatic algorithms for analysis of sequence data, implemented SMIPS for clinical validation and performed functional studies. M.P., L.C., P.J., M.J., M.L.-G., D.Colomer, S.G., C.L., A.N. and J.Y. were responsible for downstream validation analysis and functional studies. S.B., A.C., G.E., L.B., J.M.C.T., J.C.D. and C.T. studied structural variants. M.G., M.B., S.H., D.A.P., H.H., E.C., J.M., K.R. and L.A.S. were responsible for generating libraries, performing exome capture and running sequencers. P.N., C.M.R.-C. and M.A. prepared and supervised the bioethical requirements. S.S. designed the epidemiologic study. P.K. gave conceptual advice and revised the manuscript. N.V., T.B., D.Costa, A.C., A.E., L.H., M.A.P., E.A., J.S.M., E.M., A.L.-G. and E.C. performed clinical and biological studies. M.R., N.V., J.M.H. and M.G.-D. were the pathologists who reviewed and confirmed the diagnoses. R.R., J.L.G., M.O., D.T., D.G.P., A.V. and R.G. were in charge of bioinformatics data management. I.G., P.A.F., M.R.S. and P.J.C. coordinated the sequencing efforts and performed primary data analysis. X.S.P., X.E., A.L.-G., C.L.-O. and E.C. directed the research and wrote the manuscript, which all authors have approved.

Author Information Sequencing, expression and genotyping array data have been deposited at the European Genome-Phenome Archive (EGA, <http://www.ebi.ac.uk/ega/>), which is hosted at the European Bioinformatics Institute (EBI), under accession number EGAS00000000092. Reprints and permissions information is available at www.nature.com/reprints. This paper is distributed under the terms of the Creative Commons Attribution-Non-Commercial-Share Alike licence, and is freely available to all readers at www.nature.com/nature. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to E.C. (ecampo@clinic.ub.es) or C.L.-O. (clo@uniovi.es).

Oncogene-induced Nrf2 transcription promotes ROS detoxification and tumorigenesis

Gina M. DeNicola^{1,2}, Florian A. Karreth^{1,3}, Timothy J. Humpton¹, Aarthi Gopinathan^{1,2}, Cong Wei⁴, Kristopher Frese¹, Dipti Mangal⁴, Kenneth H. Yu⁴, Charles J. Yeo⁵, Eric S. Calhoun⁶, Francesca Scrimieri⁷, Jordan M. Winter⁸, Ralph H. Hruban^{7,9}, Christine Iacobuzio-Donahue^{7,9}, Scott E. Kern^{7,9}, Ian A. Blair⁴ & David A. Tuveson¹

Reactive oxygen species (ROS) are mutagenic and may thereby promote cancer¹. Normally, ROS levels are tightly controlled by an inducible antioxidant program that responds to cellular stressors and is predominantly regulated by the transcription factor Nrf2 (also known as Nfe2l2) and its repressor protein Keap1 (refs 2–5). In contrast to the acute physiological regulation of Nrf2, in neoplasia there is evidence for increased basal activation of Nrf2. Indeed, somatic mutations that disrupt the Nrf2–Keap1 interaction to stabilize Nrf2 and increase the constitutive transcription of Nrf2 target genes were recently identified, indicating that enhanced ROS detoxification and additional Nrf2 functions may in fact be protumorigenic⁶. Here, we investigated ROS metabolism in primary murine cells following the expression of endogenous oncogenic alleles of *Kras*, *Braf* and *Myc*, and found that ROS are actively suppressed by these oncogenes. K-Ras^{G12D}, B-Raf^{V619E} and Myc^{ERT2} each increased the transcription of Nrf2 to stably elevate the basal Nrf2 antioxidant program and thereby lower intracellular ROS and confer a more reduced intracellular environment. Oncogene-directed increased expression of Nrf2 is a new mechanism for the activation of the Nrf2 antioxidant program, and is evident in primary cells and tissues of mice expressing K-Ras^{G12D} and B-Raf^{V619E}, and in human pancreatic cancer. Furthermore, genetic targeting of the Nrf2 pathway impairs K-Ras^{G12D}-induced proliferation and tumorigenesis *in vivo*. Thus, the Nrf2 antioxidant and cellular detoxification program represents a previously unappreciated mediator of oncogenesis.

To examine the role of ROS in cellular transformation and tumorigenesis, we used an endogenous and conditional oncogenic *LSL-K-Ras^{G12D}* allele⁷. K-Ras^{G12D/+} and K-Ras^{LSL/+} mouse embryonic fibroblasts (MEFs) were compared to MEFs and NIH3T3 fibroblasts (NIH3T3s) transduced with *K-Ras4B^{G12D}* and *H-Ras^{V12}* (Fig. 1a), because ectopic Ras introduction was previously reported to increase ROS production⁸. In contrast to ectopic overexpression of oncogenic Ras, K-Ras^{G12D/+} MEFs demonstrated lower levels of hydrogen peroxide, superoxide and mitochondrial ROS compared to K-Ras^{LSL/+} MEFs (Fig. 1a and Supplementary Fig. 2a, b). Accordingly, the level of 7,8-dihydro-8-oxo-2'-deoxyguanosine (8-oxo-dGuo), one of the major products of DNA oxidation, was decreased in MEFs expressing endogenous K-Ras^{G12D/+}, but increased by ectopic oncogenic Ras (Fig. 1b). The levels of ROS and 8-oxo-dGuo were also lower in p53^{-/-}; K-Ras^{G12D/+} MEFs compared to p53^{-/-}; K-Ras^{LSL/+} MEFs (p53 is also known as Trp53; Supplementary Fig. 2c, d), supporting that these differences are not due to differential activation of a senescence program and demonstrating that increased ROS production is not required for full transformation^{7,9}.

Because alterations of ROS can affect the intracellular redox state¹⁰, the ratio of reduced to oxidized glutathione (GSH/GSSG) was examined in cells expressing ectopic and endogenous K-Ras^{G12D}. Whereas both

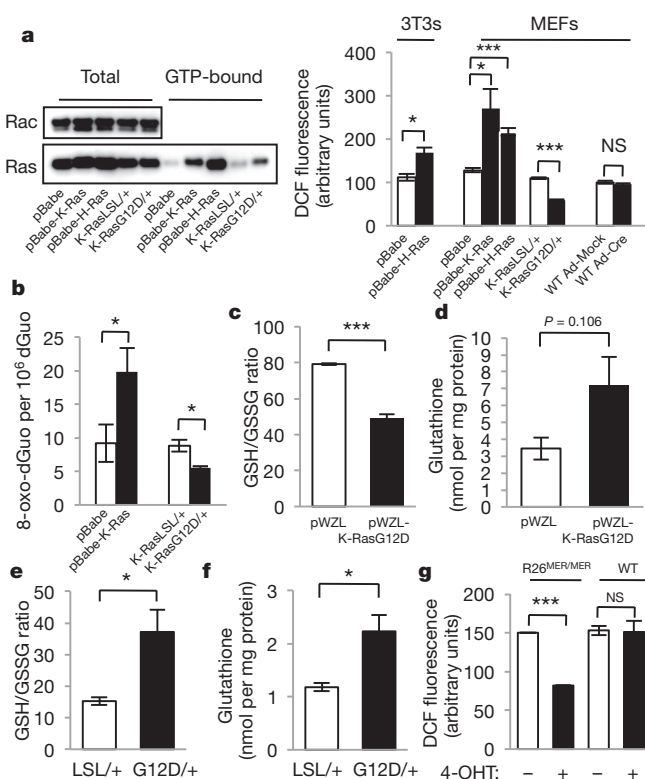


Figure 1 | Physiological expression of oncogenes lowers ROS. NIH3T3s and MEFs were transduced with retroviral vectors and evaluated 6 days later: control vector (pBabe), pBabe-H-Ras^{G12V} (p-Babe-H-Ras), or pBabe-K-Ras^{G12D} (p-Babe-K-Ras). Alternatively, LSL-K-Ras^{G12D} MEFs were infected with Ad-mock (K-Ras^{LSL/+}) or Ad-cre (K-Ras^{G12D/+}) and evaluated 4 days later. Wild-type MEFs were infected with Ad-mock (WT Ad-Mock) or Ad-cre (WT Ad-Cre) and used as controls. **a**, Left, western blot of total and GTP-bound Ras in MEFs expressing endogenous and ectopic Ras, with Rac used as a loading control. Right, ROS levels following expression of oncogenic Ras, as determined by 2',7'-dichlorofluorescein diacetate (DCF) staining. **b**, 8-oxo-dGuo levels following ectopic and endogenous expression of K-Ras^{G12D}. **c–f**, Determination of the GSH/GSSG ratios and total cellular glutathione in cells overexpressing ectopic K-Ras^{G12D} (**c**, **d**), or expressing endogenous K-Ras^{G12D} (**e**, **f**). **g**, ROS levels following activation of Myc^{ERT2}. R26^{MER/MER} MEFs were treated with DMSO or 100 nM 4-OHT and assayed after 24 h. Data are representative of 3 or more independent experiments. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$ and error bars represent \pm s.e.m. here and for all figures.

¹Li Ka Shing Centre, Cancer Research UK Cambridge Institute, Robinson Way, Cambridge CB2 0RE, UK. ²Abramson Family Cancer Research Institute, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. ³University of Vienna, Dr. Bohrgasse 9, 1030 Vienna, Austria. ⁴Center for Cancer Pharmacology, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. ⁵Department of Surgery, Jefferson Medical College, Philadelphia, Pennsylvania 19107, USA. ⁶Department of Biology, Alma College, Alma, Michigan 48801, USA. ⁷Department of Oncology, The Sol Goldman Pancreatic Cancer Research Center, the Johns Hopkins Medical Institutions, Baltimore, Maryland 21287, USA. ⁸Department of Surgery, The Sol Goldman Pancreatic Cancer Research Center, the Johns Hopkins Medical Institutions, Baltimore, Maryland 21287, USA. ⁹Department of Pathology, The Sol Goldman Pancreatic Cancer Research Center, the Johns Hopkins Medical Institutions, Baltimore, Maryland 21287, USA.

elevated the level of total glutathione (Fig. 1d, f and ref. 11), only the endogenous expression of K-Ras^{G12D} elevated the GSH/GSSG ratio to promote a more reduced intracellular environment (Fig. 1c, e). The production of ROS by ectopic oncogenic Ras is chiefly regulated by NADPH-oxidase (Nox)¹². Only the ectopic overexpression of oncogenic Ras increased Nox activity and messenger RNA, providing a mechanistic explanation for these differing results (Supplementary Fig. 3). Similar to oncogenic Ras, ectopic expression of Myc has also been shown to increase the production of ROS in NIH3T3 fibroblasts¹³ and accordingly we sought to determine whether near-physiological expression of this oncogene had a similar effect. Treatment of R26^{MER/MER} MEFs, which express a homozygous Myc^{ERT2} allele under the control of the Rosa26 promoter¹⁴, with 4-hydroxytamoxifen (4-OHT) resulted in a substantial reduction of ROS (Fig. 1g). However, this was not a characteristic feature of all oncogenic stimuli as expression of activated Notch1 and β -catenin did not lower ROS (Supplementary Fig. 4). Therefore, endogenous expression of oncogenic K-Ras or Myc^{ERT2} lowers the level of cellular ROS, in contrast to the ectopic overexpression of these oncogenes.

As the intracellular redox state is chiefly regulated by Nrf2, we examined Nrf2 protein levels and activity in K-Ras^{LSL/+} and K-Ras^{G12D/+} MEFs by western blot, reporter assay and chromatin immunoprecipitation (ChIP) analysis of the Nrf2 target genes *Hmox1* and *Nqo1*. Expression of K-Ras^{G12D} resulted in an approximately twofold increase in Nrf2 protein and binding activity compared to control MEFs (Fig. 2a, b and Supplementary Fig. 5a), which was not attributable to altered regulation by Keap1 (ref. 2 and Supplementary Fig. 6). K-Ras^{G12D/+} MEFs showed increased *Nrf2* mRNA and increased expression of *Hmox1*, *Nqo1*, *Gclc*, *Gclm* and *Ggt1* mRNA and protein (Fig. 2c and Supplementary Fig. 5b). Increased expression of Nrf2 and its target genes was also observed following K-Ras^{G12D} expression in p53^{-/-} MEFs, and following ectopic expression of K-Ras^{G12D} and H-Ras^{V12} in primary MEFs, but not following expression of activated Notch1 or β -catenin (Supplementary Fig. 5c–g). However, expression of K-Ras^{G12D} in Nrf2-deficient MEFs failed to

elevate total glutathione levels and resulted in a more oxidized intracellular environment (Fig. 2d, e). Neither the cell culture conditions employed to express K-Ras^{G12D} nor the gene dosage of wild-type K-Ras affected the expression of Nrf2 target genes (Supplementary Fig. 7a, b). Additionally, ROS metabolism in wild-type MEFs was sensitive to acute changes in the levels of Keap1 and Nrf2, further supporting a causal relationship between Nrf2 and ROS (Supplementary Fig. 7c–e). Furthermore, acute knockdown of Nrf2 attenuated the reduction in ROS by K-Ras^{G12D} (Fig. 2f), and the effects of Nrf2 depletion on ROS were dosage-dependent, supporting the importance of the level of *Nrf2* mRNA for ROS control (Supplementary Fig. 7f, g). Similar to K-Ras^{G12D}, activation of c-Myc^{ERT2} (with 4-OHT) promoted an increase in the mRNA and protein levels of Nrf2 and its target genes (Fig. 2g, h and Supplementary Fig. 7h). Furthermore, ChIP-seq data from the ENCODE consortium demonstrated direct binding of Myc to the *Nfe2l2* locus (Supplementary Fig. 8a)¹⁵. Therefore, the K-Ras and Myc oncogenes can constitutively increase the transcription of *Nrf2* to elevate the basal activity of the antioxidant and cellular detoxification program.

To investigate the mechanism of Nrf2 activation by K-Ras^{G12D}, the roles of the Raf/MEK/ERK and p38alpha (also known as Araf/Braf/Raf1, Map2k1/2, Mapk1/2 and Mapk14) MAPK pathways were investigated. First, cells were treated with a potent and specific inhibitor of MEK, AZD6244 (ARRY-142886) (Supplementary Fig. 9a–c), which restored the ROS level of K-Ras^{G12D/+} cells almost to the level of K-Ras^{LSL/+} cells (Fig. 3a). Additionally, AZD6244 treatment resulted in decreased induction of Nrf2 and its target genes (Fig. 3b). Furthermore, endogenous expression of B-Raf^{V619E} (corresponding to human B-Raf^{V600E})¹⁶ resulted in increased phospho-ERK levels, a decrease in ROS, and an increase in *Nrf2* mRNA and antioxidant gene expression (Supplementary Fig. 9d–f). As previously reported¹⁷, we found that p38alpha MAPK did not activate Nrf2 (Supplementary Fig. 9g–i). To determine the mechanism of increased Nrf2 expression, transcription factors downstream of MAPK signalling were examined. Accordingly, knockdown of Jun, Fra1 (also known as Fos1) and Myc, but not JunD or Elk1,

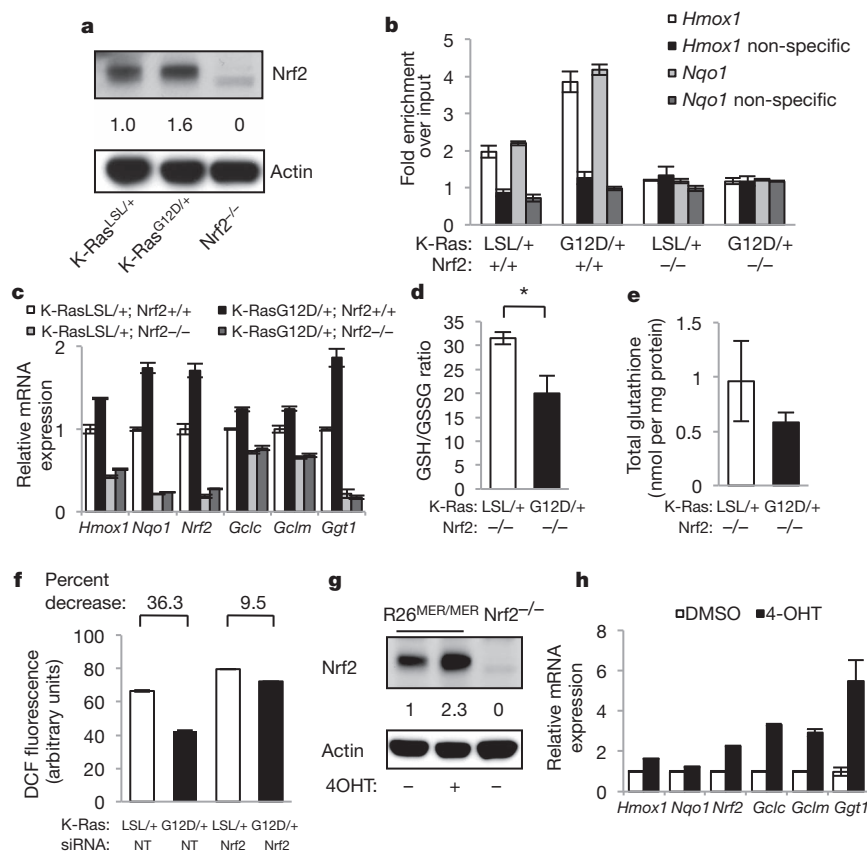


Figure 2 | Physiological expression of oncogenes activates the Nrf2 antioxidant program.

a, Western blot demonstrates a 60% increase in Nrf2 protein following expression of endogenous K-Ras^{G12D}. Antibody specificity was confirmed using Nrf2^{-/-} MEFs. **b**, Nrf2 ChIP followed by q-PCR for the *Hmox1* and *Nqo1* promoters. Control non-specific primers amplified regions of DNA located 50kb from the *Hmox1* and *Nqo1* promoters. **c**, Expression of Nrf2 and Nrf2 target genes *Nqo1*, *Hmox1*, *Gclc*, *Gclm* and *Ggt1* upon K-Ras^{G12D} expression in Nrf2^{+/+} and Nrf2^{-/-} MEFs. *Nrf2* mRNA is relatively unstable but still detectable at low levels in Nrf2^{-/-} MEFs. **d**, **e**, Determination of the GSH/GSSG ratio (**d**) and total glutathione (**e**) upon K-Ras^{G12D} expression in Nrf2^{-/-} MEFs. **f**, ROS levels following Nrf2 depletion with siRNA. LSL-K-Ras^{G12D} MEFs were transfected with non-targeting (NT) or *Nrf2* siRNA, infected with Ad-mock or Ad-cre and assayed after 48 h for DCF oxidation. **g**, Western blot of Nrf2 protein levels following induction of Myc^{ERT2} by 4-OHT. Densitometry shows a 2.3-fold increase. **h**, Analysis of Nrf2 antioxidant program gene expression following activation of Myc^{ERT2}. R26^{MER/MER} MEFs were treated with DMSO or 100 nM 4-OHT for 24 h and assayed for antioxidant gene expression. Data are representative of three independent experiments.

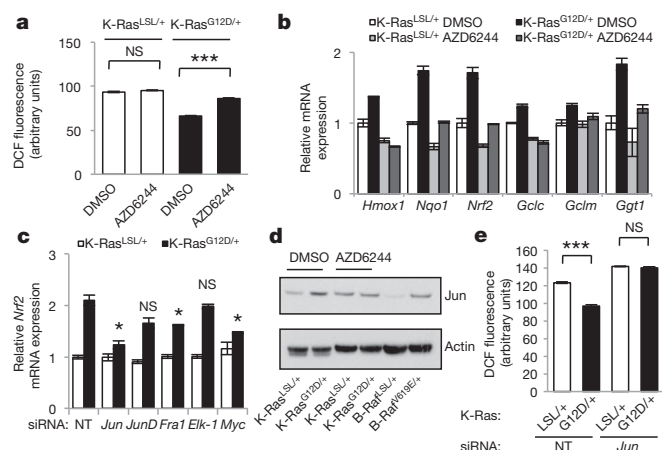


Figure 3 | Activation of Nrf2 by K-Ras^{G12D} occurs via the Raf-MEK-ERK-Jun pathway. **a**, ROS levels following treatment of K-Ras^{G12D} MEFs with AZD6244. LSL-K-Ras^{G12D} MEFs were treated with DMSO or 0.1 μ M AZD6244, infected with Ad-mock (K-Ras^{LSL/+}) or Ad-cre (K-Ras^{G12D/+}) and assayed after 72 h. **b**, Analysis of antioxidant gene expression following treatment of K-Ras^{LSL/+} and K-Ras^{G12D/+} MEFs with AZD6244 for 24 h. **c**, Control of Nrf2 transcription by AP-1 family members. K-Ras^{LSL/+} and K-Ras^{G12D/+} MEFs were transfected with siRNA and assayed for Nrf2 expression after 48 h. **d**, Western blot of Jun and actin protein levels in LSL-K-Ras^{G12D} and LSL-B-Raf^{V619E} MEFs. K-Ras^{LSL/+} and K-Ras^{G12D/+} MEFs were treated with DMSO or 0.1 μ M AZD6244 for 24 h. **e**, ROS levels following Jun depletion with siRNA. LSL-K-Ras^{G12D} MEFs were transfected with non-targeting (NT) or Jun siRNA, infected with Ad-mock or Ad-cre and assayed after 48 h for DCF oxidation. Data are representative of three independent experiments.

decreased the Nrf2 mRNA in K-Ras^{G12D/+} cells, with almost complete rescue achieved with Jun (Fig. 3c). Short interfering RNA (siRNA) efficiency was confirmed by real-time PCR and western blot (Supplementary Fig. 10a, b). Importantly, K-Ras^{G12D/+} MEFs and B-Raf^{V619E} MEFs showed elevated Jun protein levels compared to control MEFs, and the elevated Jun level in K-Ras^{G12D/+} MEFs was rescued by treatment with AZD6244 (Fig. 3d). Furthermore, depletion of Jun with siRNA prevented the decrease in ROS following expression of K-Ras^{G12D} (Fig. 3e, Supplementary Fig. 10c). Our results reinforce a prior finding that antioxidant response elements that closely resemble AP-1 sites help regulate the Nrf2 promoter¹⁸. Finally, ChIP-seq revealed that the transcriptional start site of the Nfe2l2 locus is a direct binding target of Jun (Supplementary Fig. 8b)¹⁵. Thus, K-Ras^{G12D} and B-Raf^{V619E} stimulate transcription of Nrf2 via Jun and Myc.

We next examined whether these oncogenes promote activation of the Nrf2 antioxidant and cellular detoxification program *in vivo*. Elevated protein expression of the Nrf2 target gene Nqo1 and decreased immunoreactivity for 8-oxo-dGuo were evident in K-Ras mutant murine¹⁹ and human pre-invasive pancreatic intraepithelial neoplasia (PanIN) and pancreatic ductal adenocarcinoma (PDA), and in murine lung adenomas expressing B-Raf^{V619E} (Fig. 4a and Supplementary Figs 11a, 12, 13 and 14a). PanIN also demonstrated reduced immunoreactivity for the lipid peroxidation adduct malondialdehyde (MDA) (Supplementary Fig. 11b). Furthermore, K-Ras^{G12D/+} pancreatic epithelial cells showed activation of Nrf2 and decreased ROS compared to K-Ras^{LSL/+}, whereas treatment of human PDA tumour lines with KRAS siRNA resulted in decreased Nrf2 and Nqo1 mRNA and an increase in ROS (Supplementary Fig. 15). The activation of Nrf2 in human pancreatic cancer could not be explained by somatic mutations in NFE2L2 or KEAP1; we sequenced over 100 samples of human PDA and identified only one case with a concomitant KEAP1 and KRAS mutation (Supplementary Information and Supplementary Fig. 14b). Importantly, Nrf2-deficient murine PanIN were negative for Nqo1 and demonstrated similar levels of 8-oxo-dGuo and MDA in PanIN compared to neighbouring morphologically normal ductal cells, supporting

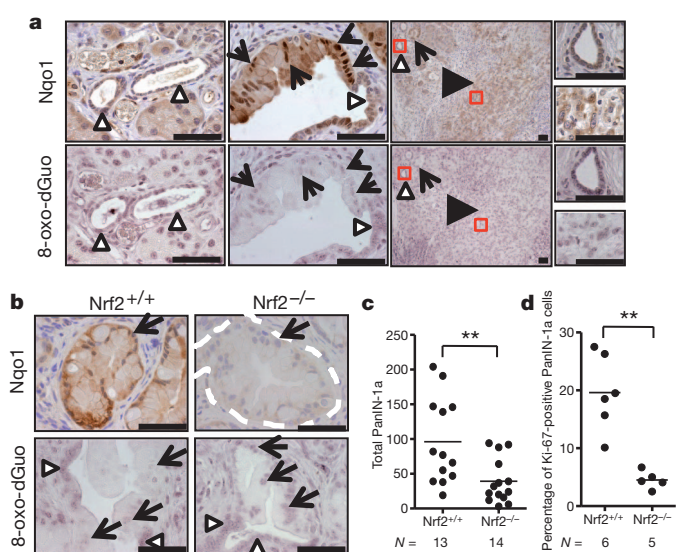


Figure 4 | Evidence for Nrf2 antioxidant program in pancreatic cancer.

a, Immunohistochemical detection of Nqo1 (brown staining) and 8-oxo-dGuo (purple staining) in mouse PanIN and PDA (similar patterns observed for 11 out of 11 of cases examined) in comparison to morphologically normal ducts. PanIN (arrows), PDA (black arrowheads), normal ducts (white arrowheads) here and for all figures. Scale bar, 56 μ m. **b**, Immunohistochemical detection of Nqo1 and 8-oxo-dGuo in Nrf2^{-/-} PanIN compared to Nrf2^{+/+} PanIN (similar patterns observed for 5 out of 5 of each genotype examined, PanIN outlined by white dashes). Scale bar = 56 μ m. **c**, Nrf2^{-/-} and Nrf2^{+/+} PanIN-1a incidence. Whole pancreata were sectioned at 100- μ m intervals and total numbers of PanIN-1a were counted. **d**, Proliferation of PanIN-1a cells in Nrf2^{-/-} and Nrf2^{+/+} mice, as determined by Ki-67 immunostaining.

a role for Nrf2 in ROS detoxification *in vivo* (Fig. 4b and Supplementary Fig. 11b, c). Nrf2-deficient salivary gland did not demonstrate elevated 8-oxo-dGuo immunoreactivity, indicating that K-Ras^{G12D}-expressing cells are more reliant on Nrf2 for ROS detoxification (Supplementary Fig. 16a). Therefore, an antioxidant program is operant during tumorigenesis, consistent with our findings that link oncogenic K-Ras and B-Raf expression with the activation of Nrf2.

To investigate whether activation of Nrf2 promotes K-Ras^{G12D}-initiated carcinomas, the effects of Nrf2 ablation were examined in established mouse models of pancreatic and lung cancer. Nrf2-deficient pancreata were noted to contain fewer PanIN (Fig. 4c). Whereas the proliferation of the salivary gland was unchanged (Supplementary Fig. 16b), PanIN from Nrf2-deficient mice were significantly less proliferative and demonstrated an increased content of cells showing senescence-associated β -galactosidase activity (Fig. 4d and Supplementary Fig. 17a). No significant differences were observed in PanIN cell apoptosis or DNA damage (Supplementary Fig. 17b, c). Senescence was also observed in K-Ras^{G12D/+} MEFs following acute depletion of Nrf2 with siRNA and was dependent on expression of p53 (Supplementary Fig. 17d). The Nrf2-dependent proliferation defect noted in PanIN was also observed in MEFs, and both could be rescued by the antioxidant N-acetyl cysteine (Supplementary Fig. 18a, b). Furthermore, treatment of Nrf2^{+/+} PanIN mice with a glutathione synthesis inhibitor resulted in a significant decrease in the proliferation of PanIN but not adjacent acinar cells (Supplementary Fig. 18c–e). Additionally, the role of Nrf2 in lung cancer development was investigated. Nrf2 deficiency resulted in a significant reduction in disease burden and proliferation, and an increase in median survival (Supplementary Fig. 19). Although we cannot exclude the potential involvement of non-cell-autonomous effects arising from constitutive Nrf2 deficiency, the major impact seemed to be restricted to tumour initiation and proliferation of nascent preneoplastic cells. These results demonstrate that Nrf2 promotes K-Ras^{G12D}-initiated pancreatic and lung tumorigenesis and proliferation.

ROS can stimulate tumorigenesis through the oxidation of DNA and subsequent mutation of genes that promote carcinogenesis. In contrast, we found that activating a ROS-detoxification program contributed to tumorigenesis, although our study does not exclude a role for other Nrf2 targets such as drug-metabolizing enzymes and efflux pumps, heat shock proteins, 26S proteasome subunits, growth factors and receptors⁶. Future studies will be necessary to determine the role of these Nrf2 functions in tumours with activated Nrf2. Increased expression of Nrf2 target genes and increased stability of Nrf2 caused by somatic mutations in Nrf2 and Keap1 are well documented in human cancer, providing one mechanism for enhanced Nrf2 activity during tumorigenesis. Our work describes oncogenic signalling as an alternative mechanism to activate *Nrf2* transcription during tumorigenesis (Supplementary Fig. 1), and suggests that modulation of the redox state is uniformly important in cancer and may therefore represent a therapeutic opportunity. Thus, constitutively elevated Nrf2 activity in cancer cells occurs through two distinct mechanisms: diminished Nrf2 turnover and augmented *Nrf2* mRNA levels.

METHODS SUMMARY

Mouse strains were described previously^{20–22}, and were interbred onto the same genetic background (mixed B6/129/SJL) for tumour studies. Mice were infected intranasally with 10⁷ plaque-forming units of adenoviral-cre as described²⁰. Mice were maintained in compliance with UK Home Office regulations. Mouse embryonic fibroblasts were isolated from embryonic day 13.5 (E13.5) embryos, maintained in DMEM with 10% FCS and were plated at 70% confluence the day before use unless otherwise stated. Early-passage MEFs (passage 3–4) were used for all experiments, and at least three lines were examined for all studies. R26^{MER/MER} (G. Evan), R26-LSL-NIC (C. Murtaugh), APC^{fl/fl} (O. Sansom), p38a^{-/-}; LSL-K-Ras^{G12D} (J.-J. Ventura), and Jun^{-/-} (E. Wagner) MEFs were generously provided. Adenoviral and retroviral transductions were performed as described⁷. NIH3T3 fibroblasts were obtained from the ATCC and maintained in DMEM with 10% FCS. For ROS measurements, cells were labelled in serum- and phenol-red-free media containing 5 µM dihydroethidium or 10 µM 2',7'-dichlorofluorescein diacetate (both Sigma Aldrich) for 30 min, and analysed by flow cytometry. 8-oxo-dGuo and glutathione were measured by liquid chromatography-mass spectrometry as described^{23,24}. Protein lysates were prepared using 1% SDS lysis buffer and separated on 4–12% NuPAGE gels (Invitrogen), transferred onto a PVDF membrane (Millipore) and incubated with the following antibodies: actin (I-19), Keap1 (E-20), Nqo1 (C-19), Ggt1 (3E6) and B-Raf (F-7) (all Santa Cruz Biotechnologies); ERK, phospho-ERK (Thr202/Tyr204), Akt, phospho-AKT (S473), Jun and Elk1 (all Cell Signaling); Hmox1 (Stressgen) and β-catenin (BD Biosciences) were also used. Nrf2 western blots are described in the supplementary methods. RNA was isolated using an RNeasy kit (Qiagen). cDNA was synthesized using a High Capacity RNA to cDNA Master Mix and analysed by quantitative PCR (q-PCR) using TaqMan gene expression assays on a 7900HT Real-Time PCR system (all Applied Biosystems). Target gene expression was normalized to actin expression, and shown relative to control samples.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 8 August 2010; accepted 12 May 2011.

- Shibutani, S., Takeshita, M. & Grollman, A. P. Insertion of specific bases during DNA synthesis past the oxidation-damaged base 8-oxodG. *Nature* **349**, 431–434 (1991).
- Itoh, K. *et al.* Keap1 represses nuclear activation of antioxidant responsive elements by Nrf2 through binding to the amino-terminal Neh2 domain. *Genes Dev.* **13**, 76–86 (1999).
- Wakabayashi, N. *et al.* Keap1-null mutation leads to postnatal lethality due to constitutive Nrf2 activation. *Nature Genet.* **35**, 238–245 (2003).
- Nguyen, T., Nioi, P. & Pickett, C. B. The Nrf2-antioxidant response element signaling pathway and its activation by oxidative stress. *J. Biol. Chem.* **284**, 13291–13295 (2009).
- Venugopal, R. & Jaiswal, A. K. Nrf2 and Nrf1 in association with Jun proteins regulate antioxidant response element-mediated expression and coordinated induction of genes encoding detoxifying enzymes. *Oncogene* **17**, 3145–3156 (1998).

- Hayes, J. D. & McMahon, M. *NRF2* and *KEAP1* mutations: permanent activation of an adaptive response in cancer. *Trends Biochem. Sci.* **34**, 176–188 (2009).
- Tuveson, D. A. *et al.* Endogenous oncogenic *K-ras*^{G12D} stimulates proliferation and widespread neoplastic and developmental defects. *Cancer Cell* **5**, 375–387 (2004).
- Lee, A. C. *et al.* Ras proteins induce senescence by altering the intracellular levels of reactive oxygen species. *J. Biol. Chem.* **274**, 7936–7940 (1999).
- Mitsushita, J., Lambeth, J. D. & Kamata, T. The superoxide-generating oxidase Nox1 is functionally required for Ras oncogene transformation. *Cancer Res.* **64**, 3580–3585 (2004).
- Cakir, Y. & Ballinger, S. W. Reactive species-mediated regulation of cell signaling and the cell cycle: the role of MAPK. *Antioxid. Redox Signal.* **7**, 726–740 (2005).
- Recktenwald, C. V., Kellner, R., Lichtenfels, R. & Seliger, B. Altered detoxification status and increased resistance to oxidative stress by *K-ras* transformation. *Cancer Res.* **68**, 10086–10093 (2008).
- Irani, K. *et al.* Mitogenic signaling mediated by oxidants in Ras-transformed fibroblasts. *Science* **275**, 1649–1652 (1997).
- Tanaka, H. *et al.* E2F1 and c-Myc potentiate apoptosis through inhibition of NF-κB activity that facilitates MnSOD-mediated ROS elimination. *Mol. Cell* **9**, 1017–1029 (2002).
- Murphy, D. J. *et al.* Distinct thresholds govern Myc's biological output *in vivo*. *Cancer Cell* **14**, 447–457 (2008).
- Birney, E. *et al.* Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799–816 (2007).
- Karreth, F. A., DeNicola, G. M., Winter, S. P. & Tuveson, D. A. C-Raf inhibits MAPK activation and transformation by B-Raf^{V600E}. *Mol. Cell* **36**, 477–486 (2009).
- Keum, Y. S. *et al.* Mechanism of action of sulforaphane: inhibition of p38 mitogen-activated protein kinase isoforms contributing to the induction of antioxidant response element-mediated heme oxygenase-1 in human hepatoma HepG2 cells. *Cancer Res.* **66**, 8804–8813 (2006).
- Kwak, M. K., Itoh, K., Yamamoto, M. & Kensler, T. W. Enhanced expression of the transcription factor Nrf2 by cancer chemopreventive agents: role of antioxidant response element-like sequences in the *nrf2* promoter. *Mol. Cell. Biol.* **22**, 2883–2892 (2002).
- Hingorani, S. R. *et al.* Preinvasive and invasive ductal pancreatic cancer and its early detection in the mouse. *Cancer Cell* **4**, 437–450 (2003).
- Jackson, E. L. *et al.* Analysis of lung tumor initiation and progression using conditional expression of oncogenic *K-ras*. *Genes Dev.* **15**, 3243–3248 (2001).
- Chan, K., Lu, R., Chang, J. C. & Kan, Y. W. NFE2, a member of the NFE2 family of transcription factors, is not essential for murine erythropoiesis, growth, and development. *Proc. Natl Acad. Sci. USA* **93**, 13943–13948 (1996).
- Kawaguchi, Y. *et al.* The role of the transcriptional regulator Ptf1a in converting intestinal to pancreatic progenitors. *Nature Genet.* **32**, 128–134 (2002).
- Park, J. H. *et al.* Evidence for the aldo-keto reductase pathway of polycyclic aromatic trans-dihydrodiol activation in human lung A549 cells. *Proc. Natl Acad. Sci. USA* **105**, 6846–6851 (2008).
- Zhu, P., Oe, T. & Blair, I. A. Determination of cellular redox status by stable isotope dilution liquid chromatography/mass spectrometry analysis of glutathione and glutathione disulfide. *Rapid Commun. Mass Spectrom.* **22**, 432–440 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Johnson for providing the Nrf2^{-/-} mice; G. Evan, O. Sansom, C. Murtaugh, J.-J. Ventura and E. Wagner for MEFs; E. Schmidt for Nrf2 antiserum; E. Jaffee, A. Maitra and A. Horii for human PDA cell lines; B. Haynes, S. Davies and N. Cook for human PDA tissue samples; C. Ross-Innes, K. Holmes and J. Carroll for advice with the ChIP assay; and the ENCODE Consortium for ChIP-seq studies. We thank F. Connor, C. Martins and other members of the Tuveson lab for assistance and advice, and the animal care staff and histology core at CRI. This research was supported by the University of Cambridge and Cancer Research UK, The Li Ka Shing Foundation and Hutchison Whampoa Limited, the NIHR Cambridge Biomedical Research Centre, and the NIH (grants CA101973, CA111294, CA084291 and CA105490 to D.A.T.; CA62924 and CA128920 to S.E.K. and C.I.-D.; and CA106610 to C.I.-D.). Additional support was obtained from the Abramson Cancer Center of the University of Pennsylvania Pilot Grant IRG 78-002-26 (D.A.T.), Emerald Foundation (E.S.C.), the Marjorie Kovler Fund (S.E.K.) and the Ruth L. Kirschstein National Research Service Award F32CA123887-01 (K.F.). We regret that many primary references have been omitted due to space limitations.

Author Contributions G.M.D., F.A.K., T.J.H., A.G. and K.F. performed cell culture and mouse experiments. C.W., D.M., K.H.Y. and I.A.B. performed 8-oxo-dGuo and glutathione assays. C.J.Y., E.S.C., F.S., J.M.W., R.H.H., C.I.-D. and S.E.K. performed Nrf2 and Keap1 sequencing. G.M.D. and D.A.T. designed the study and wrote the manuscript, and all authors commented on it.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to D.A.T. (david.tuveson@cancer.org.uk).

METHODS

Mouse strains and MEFs. Mouse strains were described previously^{20–22}, and were interbred onto the same genetic background (mixed B6/129/SJL) for tumour studies. Mice were infected intranasally with 10⁷ plaque-forming units of adenoviral-cre as described²⁰. Mice were maintained in compliance with UK Home Office regulations. Mouse embryonic fibroblasts were isolated from embryonic day 13.5 (E13.5) embryos, maintained in DMEM with 10% FCS and were plated at 70% confluence the day before use unless otherwise stated. Early-passage MEFs (passage 3–4) were used for all experiments, and at least three lines were examined for all studies. R26^{MER/MER} (G. Evan), R26-LSL-NIC (C. Murtaugh), APC^{fl/fl} (O. Sansom), p38a^{-/-}; LSL-K-Ras^{G12D} (J.-J. Ventura) and Jun^{-/-} (E. Wagner) MEFs were provided. Adenoviral and retroviral transductions were performed as described⁷. NIH3T3 fibroblasts were obtained from the ATCC and maintained in DMEM with 10% FCS.

Analysis of 8-oxo-dGuo and glutathione. 8-oxo-dGuo and glutathione were measured by liquid chromatography-mass spectrometry as described^{23,24}.

Chromatin immunoprecipitation. Chromatin immunoprecipitation was performed as described²⁵. Cells (5 × 10⁶) were fixed at 37 °C in DMEM with 1% formaldehyde for 10 min and lysed in 1% SDS, 10 mM EDTA, 50 mM Tris-HCl pH 8.1 plus protease inhibitors and sonicated for 20 min in a cold water bath until DNA was an average size of 1 kilobase. Input was saved and lysate was diluted in immunoprecipitation buffer (1% Triton, 2 mM EDTA, 150 mM NaCl, 20 mM Tris-HCl pH 8.1) and mixed with beads (Dynal Protein A, Invitrogen) that were pre-bound overnight with Nrf2 antibody (H-300, Santa Cruz). Chromatin was immunoprecipitated overnight, and beads were washed six times with RIPA buffer (50 mM HEPES pH 7.6, 1 mM EDTA, 0.7% Na deoxycholate, 1% NP-40, 0.5 M LiCl) and twice with TE. Beads were incubated with 1% SDS, 0.1 M NaHCO₃ for 30 min at room temperature, and then crosslinks were reversed on both the input and the immunoprecipitate by heating overnight in a 65 °C water bath. DNA was purified with a QIAquick spin kit (Qiagen) and Q-PCR was performed in triplicate with a Power SYBR Master Mix (Applied Biosystems) on an Applied Biosystems 7900HT with the following primers^{26,27}: Nqo1-F, 5'-GCAGTTTCTAAGA GCAGAAC-3'; Nqo1-R, 5'-GTAGATTAGTCTCACTCAGCCG-3'; Nqo1 Non-specific-F, 5'-AGGAGATGGAAGGCAGGAAG-3'; Nqo1 Non-specific-R, 5'-GGGCGCACTATTGTCATCTT-3'; Hmox-1-F, 5'-GGGCTAGCATGCG AAGTGAG-3'; Hmox-1-R, 5'-AGACTCCGCCCTAAGGGTTC-3'; Hmox-1 Non-specific-F, 5'-GGCAGGTATGGACCTTCAAA-3'; Hmox-1 Non-specific-R, 5'-AAAGGAGTCAGGAGGGAGA-3'.

ChIP-seq (ENCODE). These data were generated and analysed by the labs of M. Snyder, M. Gerstein and S. Weissman, P. Farnham, and K. Struhl at Harvard using the K562 cell line. The following antibodies were used: c-Jun (sc-1694, Santa Cruz), c-Myc (sc-764, Santa Cruz). The data can be accessed at <http://genome.ucsc.edu/ENCODE/>.

Human pancreatic cancer cell culture. MiaPaCa2 and Hs766T cells were obtained from CRUK Clare Hall Laboratories. Both lines were grown in DMEM with 10% FBS. siRNA transfections were performed with DharmaFECT 1.

Immunohistochemistry. Human tissue was donated for research purposes by patients undergoing pancreatic surgery at a tertiary hepatopancreaticobiliary referral centre (Addenbrookes Teaching Hospitals NHS Trust, Cambridge, UK). Ethical approval was granted by the local research and ethics committee (LREC number 08/H0306/32). Paraffin-embedded mouse and human specimens were deparaffinized and rehydrated, followed by antigen retrieval in boiling citrate buffer. The following primary antibodies were used: Ki-67 (SP-6, Neomarkers), Nqo1 (Abcam, Sigma Aldrich), 8-oxo-dGuo ([N45.1], Abcam), γ -H2AX (Millipore), cleaved caspase 3 (Cell Signaling), anti-Malondialdehyde (Abcam). Horseradish-peroxidase-based detection reagents were used for all immunohistochemistry except 8-oxo-dGuo, for which alkaline-phosphatase-based reagents were used. 8-oxo-dGuo specificity was verified by treating sections for 1 h with 1 U ml⁻¹ DNase at 37 °C. For quantification of 8-oxo-dGuo, mean signal intensity was calculated using Adobe Photoshop CS. Immunohistochemistry images were converted to greyscale images, inverted, and mean pixel 8-oxo-dGuo intensity analysed. For murine Nqo1 immunohistochemistry, the Sigma Aldrich anti-Nqo1 antibody was used, and the staining pattern was confirmed to be reproducible with the Abcam antibody (not shown). The Abcam Nqo1 antibody was used for human tissue.

Luciferase assay. LSL-K-Ras^{G12D} and Nrf2^{-/-}; K-Ras^{G12D} MEFs were plated in 12-well dishes at a density of 5 × 10⁴ cells per well and infected with a Cignal Lenti ARE Reporter (Qiagen) at a multiplicity of infection of 10. After 6 h the cells were split 1:2 and infected the next day with Ad-mock and Ad-cre. After 4 days the cells were plated at 70% confluence and analysed for luciferase activity using the Dual-Luciferase Reporter Assay System (Promega) with a luminometer. Lysates were analysed in duplicate and cells lacking the ARE reporter were included as a control.

Mitochondrial and cellular ROS. For mitochondrial ROS, cells were incubated in serum- and phenol red-free DMEM with 5 μ M MitoSOX or 100 nM MitoTracker Green FM (both Invitrogen) for 30 min and analysed by flow cytometry. Mitochondrial ROS was plotted as the mitochondrial ROS signal (MitoSOX) normalized to mitochondrial mass (MitoTracker Green FM). For cellular ROS, cells were labelled in serum- and phenol-red-free media containing 5 μ M dihydroethidium or 10 μ M 2',7'-dichlorofluorescein diacetate (both Sigma Aldrich) for 30 min, and analysed by flow cytometry.

Nox activity assay. Cells were washed twice with ice-cold PBS, scraped in PBS, and centrifuged to obtain a cell pellet. The pellet was Dounce-homogenized in PBS with 1 mM MgCl₂, 1 mM EGTA and protease inhibitors and 100 μ l of homogenate was added to 100 μ l of 2× reaction buffer (PBS with 1 mM MgCl₂, 1 mM EGTA, 300 mM sucrose, 10 μ M lucigenin (Sigma Aldrich), 200 μ M NADPH (Sigma Aldrich)) and relative light units were measured with a luminometer every minute for 10 min. Specificity was confirmed by adding the specific Nox inhibitor diphenylene iodonium (10 μ M, Sigma Aldrich). Graphs show one representative measurement in the linear range (2–3 min) and are expressed as relative light units per μ g protein.

Nrf2 stability. MEFs were treated with 25 μ g ml⁻¹ cycloheximide (Acros Organics) for indicated time points and total cell lysates extracted and separated by SDS-PAGE. Gels were transferred to nitrocellulose (Biorad), and membranes were probed with anti-Nrf2 antibody, affinity-purified as described²⁸.

Quantification of proliferation and neoplasms. Values for proliferation were obtained by counting at least five fields containing more than 50 cells. For PanIN studies, only cells contained in PanIN-1a were included in the analysis (surrounding stromal, acinar or immune cells were all excluded). Proliferation is determined as a percentage and therefore represents the number of Ki-67-positive PanIN-1a cells divide by the total of PanIN-1a cells. For quantification of lung disease burden, total lung areas and neoplasm areas were calculated using Image J²⁹ and disease burden was represented as a percent of total area. For characterization of individual neoplasms, data was represented as neoplasms per lung area. For quantification of PanIN per mouse, paraffin embedded pancreata were sectioned at 100- μ m intervals and individual PanIN were counted.

Ras-GTP activity assays. Ras-GTP levels were determined according to the manufacturer's instructions (Millipore). Membranes were also blotted with anti-Rac antibody (Millipore) as a loading control.

Reagents. 4-hydroxy-tamoxifen (4-OHT) was obtained from Sigma Aldrich. AZD6244 (ARRY-142886) was obtained from Symansis.

Real-time PCR. RNA was isolated using an RNeasy kit (Qiagen). cDNA was synthesized using a High Capacity RNA to cDNA Master Mix and analysed by quantitative PCR (q-PCR) using TaqMan gene expression assays on a 7900HT Real-Time PCR system (all Applied Biosystems). Target gene expression was normalized to actin expression, and shown relative to control samples.

Senescence-associated β -galactosidase staining. Cells were fixed in 2% formaldehyde/0.2% glutaraldehyde in PBS for 5 min and stained in staining solution (citric acid/phosphate buffer pH 6.0, 150 mM NaCl, 2 mM MgCl₂, 5 mM potassium ferricyanide, 5 mM potassium ferrocyanide, and 1 mg ml⁻¹ X-gal) overnight at 37 °C. Fields of at least 50 cells were counted, in triplicate. For tissue, fresh pancreas from mice harbouring PanIN was cut into small pieces (approximately 1–3-mm cubes) and fixed for 2 h in 4% PFA in PBS on ice, and incubated in staining solution overnight at 37 °C. The pieces were then fixed in formalin overnight, followed by a quick processing and embedding in paraffin wax. Sections were cut with a thickness of 10 μ m, dewaxed with minimal exposure to xylene, and counterstained with nuclear fast red. Five fields of at least 50 cells were counted per mouse.

Sequencing of KEAP1 and NRF2 in patient samples. Xenograft-enriched pancreatic cancers were created as described previously³⁰. Cell lines were obtained from the ATCC (Aspc-1, CAPAN1, CFPAC1, Hs766T, MiaPACA2, Panc-1 and Su86.86) or from other sources. PK8 and PK9 were provided by A. Hori³¹. PL11, PL19, PL23, (provided by E. Jaffee), and XPA1 (provided by A. Maitra) were also used. Genomic DNA from immortalized peripheral blood lymphocytes of 29 patients with familial pancreatic cancer were obtained from the National Familial Pancreatic Tumor Registry (NFPT). The acquisition of human tissues from resection specimens or the NFPT was approved by the Institutional Review Board at the Johns Hopkins Hospital. The coding sequences of exons 2–6, encompassing the entire coding sequence of KEAP1, and exon 2 of NRF2 (official current gene name, NFE2L2), encompassing the KEAP1 binding domain, were amplified from genomic DNAs using intronic primers flanking each exon. Primers were designed using the Primer3 online software (<http://frodo.wi.mit.edu/primer3>). PCR-amplified products were sequenced using nested primers and an ABI Prism model 3700 Applied Biosystems). Sequence analysis used Sequencher version 4.8 software (Gene Codes). Identified variants were verified by independent PCR amplification and reverse sequencing of the amplified products. The somatic or germline nature of each variant was determined by comparing their respective

sequences to matched normal DNA available from the same patients. The following primers were used: *Keap1* Exon 2 Forward, 5'-ATCAGGTCGGG AAGTTTG-3'; Reverse, 5'-AGCCCAGAACCTCCTTTTC-3'; *Keap1* Exon 3 Forward, 5'-GTCAGCGGCAGTGATAAGTTAC-3'; Reverse, 5'-TGACAGTCC CCTAAGCATTTC-3'; *Keap1* Exon 4 Forward, 5'-TCCACGAAGGTCAGC TATAATG-3'; Reverse, 5'-TCCAGGGCTTCTGTGGTTAC-3'; *Keap1* Exon 5, 5'-TCTCTCCCCGCTTCATTTC-3'; Reverse, 5'-GCAAAAGCAGTCCACAAA AG-3'; *Keap1* Exon 6 Forward, 5'-GACCATCCCTTCTGTTCTTC-3'; Reverse, 5'-GCTTTGGACTTCTTTTGAGATG-3'; *Nrf2* Exon 2 Forward, 5'-CCACC ATCAACAGTGGCATA-3'; Reverse, 5'-AAGGCAAAGCTGGAACCTCAA-3'.

Statistical analyses. All data are expressed as the mean \pm s.e.m. Results are representative examples of three or more individual experiments. Statistical analyses were performed with the Mann-Whitney U test or Student's *t*-test (**P* < 0.05; ***P* < 0.01; ****P* < 0.001; ns, not significant). Graph Pad Prism was used for both analyses.

Transfections. DNA transfections were performed with FuGENE 6 (Roche) or Lipofectamine 2000 (Invitrogen). siRNA transfections were performed with ON-TARGETplus SMARTpool siRNA using DharmaFECT 1 reagent (both Dharmacon) according to the manufacturer's instructions. The following siRNA sequences were used. Non-targeting: 5'-UGGUUUACAUGUCGA CUA-3', 5'-UGGUUUACAUGUUGUGUGA-3', 5'-UGGUUUACAUGUUU UCUGA-3' and 5'-UGGUUUACAUGUUUCCUA-3'; *Elk1*: 5'-GGAAUGA AUACAUGCGCUC-3', 5'-CCAAGGUGGCUUAGCACGA-3', 5'-GGGAUG GUGGUGAGUUCUA-3' and 5'-ACCAAAGGGUGCAGGAAUG-3'; *Fra1*: 5'-GAACCGGAAGCACUGCAUA-3', 5'-AGGCGGAGACCGACAAA-3', 5'-GAACCUUGCUCCGCCUC-3' and 5'-GCUAAGUGCAGAAACCGAA-3'; *Jun*: 5'-CCAAGAACGUGACCGACGA-3', 5'-GCAGAGAGGAAGCGCAUGA-3', 5'-GAAACGACUUCUACGACG-3' and 5'-GAACAGGUGGCACAGCUUA-3'; *JunD*: 5'-GAACAAACGUUGGUUGCGU-3', 5'-AGCGCAAGCUGGAGCGUAU-3', 5'-CCACAUCCUGUUCGUAA-3' and 5'-AAGUCUUCGUUACGCCAAA-3'; *Keap1*: 5'-GCGCCAAUGUUGACACGGA-3', 5'-GAUAUGAGCCAGAGCGGA-3', 5'-GGAUGAUCACACCGAUGAA-3' and 5'-GUUCGAGCCUGCAGCGACU-3';

K-Ras (mouse): 5'-GAACAGUAGACACGAAACA-3', 5'-AGCAAGGAGUU ACGGGAUU-3', 5'-GGUUGGAGCUGGUGGCGUA-3' and 5'-GGUGUACA GUUAUGUGAAU-3'; *K-Ras* (human): 5'-GGAGGGCUUUCUUUGUGUA-3', 5'-UCAAAAGACAAAGUGUGUAA-3', 5'-GAAGUUAUGGAAUCCUUU-3' and 5'-GAGAUAAACACGAUGCGUAU-3'; *Myc*: 5'-GAAACGACGAGAACA GUUG-3', 5'-CCACUCACCAGCACAAACUA-3', 5'-GGACACACAACGUCU UGGA-3' and 5'-UCGAAACUCUGGUGCAUA-3'; *Nrf2*: 5'-ACUCAAUUC CCACCUAAA-3', 5'-UGGAGUAAAGUCGAGAAGUG-3', 5'-CAUGUUACG UGAUGAGGAU-3' and 5'-GGACAGCAAUUACCAUUUU-3'.

Western blotting. Protein lysates were prepared using 1% SDS lysis buffer and separated on 4–12% NuPAGE gels (Invitrogen), transferred onto a PVDF membrane (Millipore) and incubated with the following antibodies: actin (I-19), Keap1 (E-20), Nqo1 (C-19), Ggt1 (3E6) and B-Raf (F-7) (all Santa Cruz Biotechnologies); ERK, phospho-ERK (Thr202/Tyr204), Akt, phospho-AKT (S473), Jun and Elk1 (all Cell Signaling); Hmox1 (Stressgen) and β -catenin (BD Biosciences) were also used.

25. Carroll, J. S. *et al.* Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. *Cell* **122**, 33–43 (2005).
26. He, X., Chen, M. G., Lin, G. X. & Ma, Q. Arsenic induces NAD(P)H-quinone oxidoreductase I by disrupting the Nrf2-Keap1-Cul3 complex and recruiting Nrf2-Maf to the antioxidant response element enhancer. *J. Biol. Chem.* **281**, 23620–23631 (2006).
27. Sun, J. *et al.* Hemoprotein Bach1 regulates enhancer availability of heme oxygenase-1 gene. *EMBO J.* **21**, 5216–5224 (2002).
28. Suvorova, E. S. *et al.* Cytoprotective Nrf2 pathway is induced in chronically bxnrd 1-deficient hepatocytes. *PLoS ONE* **4**, e6158 (2009).
29. Rasband, W. S. ImageJ. (U. S. National Institutes of Health, 1997–2011) (<http://imagej.nih.gov/ij/>).
30. Hahn, S. A. *et al.* Allelotype of pancreatic adenocarcinoma using xenograft enrichment. *Cancer Res.* **55**, 4670–4675 (1995).
31. Sun, C. *et al.* Characterization of the mutations of the K-ras, p53, p16, and SMAD4 genes in 15 human pancreatic cancer cell lines. *Oncol. Rep.* **8**, 89–92 (2001).

Intravenous gammaglobulin suppresses inflammation through a novel T_H2 pathway

Robert M. Anthony^{1*}, Toshihiko Kobayashi^{1*}, Fredrik Wermeling¹ & Jeffrey V. Ravetch¹

High-dose intravenous immunoglobulin is a widely used therapeutic preparation of highly purified immunoglobulin G (IgG) antibodies. It is administered at high doses (1–2 grams per kilogram) for the suppression of autoantibody-triggered inflammation in a variety of clinical settings¹. This anti-inflammatory activity of intravenous immunoglobulin is triggered by a minor population of IgG crystallizable fragments (Fc), with glycans terminating in α 2,6 sialic acids (sFc) that target myeloid regulatory cells expressing the lectin dendritic-cell-specific ICAM-3 grabbing non-integrin (DC-SIGN; also known as CD209)^{2–4}. Here, to characterize this response in detail, we generated humanized DC-SIGN mice (hDC-SIGN), and demonstrate that the anti-inflammatory activity of intravenous immunoglobulin can be recapitulated by the transfer of bone-marrow-derived sFc-treated hDC-SIGN⁺ macrophages or dendritic cells into naive recipients. Furthermore, sFc administration results in the production of IL-33, which, in turn, induces expansion of IL-4-producing basophils that promote increased expression of the inhibitory Fc receptor Fc γ RIIB on effector macrophages. Systemic administration of the T_H2 cytokines IL-33 or IL-4 upregulates Fc γ RIIB on macrophages, and suppresses serum-induced arthritis. Consistent with these results, transfer of IL-33-treated basophils suppressed induced arthritic inflammation. This novel DC-SIGN- T_H2 pathway initiated by an endogenous ligand, sFc, provides an intrinsic mechanism for maintaining immune homeostasis that could be manipulated to provide therapeutic benefit in autoimmune diseases.

Binding of intravenous immunoglobulin (IVIG) or sFc to specific ICAM-3 grabbing non-integrin-related 1 (SIGN-R1) on splenic marginal zone macrophages suppresses autoantibody-mediated inflammation⁴. Although the human orthologue of SIGN-R1, DC-SIGN, showed similar binding specificity for sFc as mouse SIGN-R1, its expression pattern is broader, as it is detected systemically on myeloid-derived cells, including dendritic cells, macrophages and some monocytes^{5,6}. DC-SIGN recognizes high-mannose glycans from a variety of pathogens, and acts as a pattern recognition receptor bridging innate and adaptive immunity⁷. Ligation of DC-SIGN by bacteria-derived mannosylated glycans can induce their internalization, and also synergize with other innate receptor pathways promoting inflammation and resistance to infection. In contrast, binding of sFc to DC-SIGN requires both carbohydrate and protein determinants, and results in an anti-inflammatory response^{2,4}. The immunosuppressive potential of DC-SIGN has been documented following ligation by HIV-derived gp120 or anti-DC-SIGN antibody, which promotes the development of tolerogenic, IL-10-producing dendritic cells, and interferes with Toll-like receptor (TLR) signalling^{8,9}.

To study human DC-SIGN in the context of IVIG anti-inflammatory activity, we expressed hDC-SIGN—driven by its endogenous promoter to reproduce the characteristically broad *in vivo* expression pattern of hDC-SIGN—in a mouse. Human bacterial artificial chromosome (BAC) clones encoding the DC-SIGN gene and its regulatory regions were introduced as a transgene into mice (Fig. 1a). Transgenic mice

showed surface expression of this human lectin on dendritic cells, macrophages, and monocytes, in the peripheral blood, bone marrow and spleen, resembling the human expression pattern of DC-SIGN (Supplementary Fig. 2a–c), although a higher percentage of murine monocytes were found to express DC-SIGN.

To determine if hDC-SIGN could substitute for SIGN-R1 in mediating IVIG protection, hDC-SIGN⁺ mice were crossed to SIGN-R1-deficient animals (hDC-SIGN⁺/SIGN-R1^{−/−}) and challenged with arthritogenic K/BxN serum¹⁰. Both induction of arthritis and responsiveness to IVIG and sFc were similar in wild-type mice and hDC-SIGN⁺/SIGN-R1^{−/−} mice (Fig. 1b and Supplementary Fig. 3a). In contrast, induced arthritis was not suppressed by IVIG or sFc in

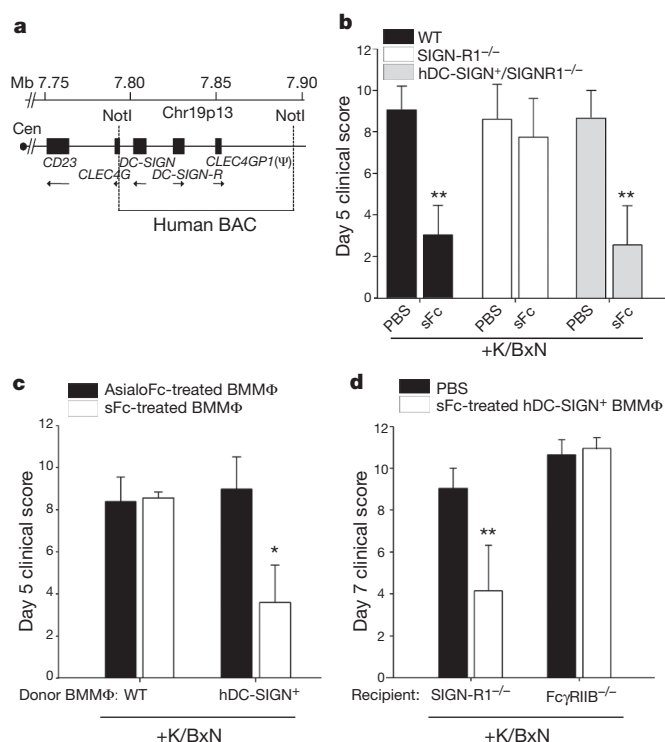


Figure 1 | Human DC-SIGN conveys sFc anti-inflammatory activity. **a**, A map of the human chromosome 19 BAC clone containing DC-SIGN and DC-SIGN-R genes. Cen, centromere. **b**, Wild-type (WT; black bars), SIGN-R1^{−/−} (white bars), or hDC-SIGN⁺/SIGN-R1^{−/−} (grey bars) mice were administered K/BxN sera and sFc. ** $P < 0.001$ determined by a Fisher least significant difference (LSD) post-hoc test. **c**, AsialoFc- (black bars) or sFc- (white bars) treated wild-type and hDC-SIGN⁺ bone-marrow derived macrophages (BMMφ) were administered to K/BxN-sera-treated wild-type recipients. * $P < 0.05$ determined by Tukey's post-hoc test. **d**, PBS (black bars) or sFc-treated hDC-SIGN⁺ bone-marrow-derived macrophages (white bars) were administered to SIGN-R1^{−/−} and Fc γ RIIB^{−/−} recipients. Means and standard deviations are plotted; ** $P < 0.001$ determined by Tukey's post-hoc test.

¹Laboratory of Molecular Genetics and Immunology, The Rockefeller University, 1230 York Avenue, New York, New York 10065, USA.

*These authors contributed equally to this work.

SIGN-R1^{-/-} mice. Thus, hDC-SIGN expression was sufficient to trigger the IVIG and sFc anti-inflammatory response.

A related lectin, DC-SIGN-R, is linked to DC-SIGN on the BAC transgene (Fig. 1a). hDC-SIGN-R has reduced affinity to sFc as compared to hDC-SIGN (Supplementary Fig. 3b). To define the contribution of DC-SIGN-R to sFc anti-inflammatory activity, mice that express hDC-SIGN alone as a transgene¹¹ were crossed with SIGN-R1^{-/-} mice (CD11c-DC-SIGN/SIGN-R1^{-/-}). These mice were protected from inflammatory arthritis by IVIG (Supplementary Fig. 3c). Further, selective blockade of hDC-SIGN in transgenic hDC-SIGN^{+/+}/SIGN-R1^{-/-} mice expressing both hDC-SIGN and hDC-SIGN-R resulted in a loss of IVIG protection *in vivo* (Supplementary Fig. 3d). These results support a requirement for hDC-SIGN but not hDC-SIGN-R in this anti-inflammatory response triggered by sFc.

Next, we sought to determine if stimulation of hDC-SIGN⁺ cells matured from bone marrow with sFc was sufficient to induce an anti-inflammatory response. Bone-marrow-derived macrophages and dendritic cells cultured from hDC-SIGN⁺ transgenic animals expressed hDC-SIGN, but not hDC-SIGN-R or SIGN-R1 (Supplementary Fig. 4a, b, c). Bone-marrow-derived cells cultured from hDC-SIGN⁺ transgenic or wild-type mice were pulsed for 30 min with sFc or asialylated Fcs (asialoFc) at a concentration representative of *in vivo* treatments. The treated cells were collected, washed and administered to wild-type mice, which were then challenged with K/BxN serum (Supplementary Fig. 4d). Mice receiving hDC-SIGN⁺ bone-marrow-derived macrophages or dendritic cells pulsed with sFc or IVIG showed reduced joint inflammation as compared to recipient mice that received wild-type cells, or hDC-SIGN⁺ cells pulsed with asialoFc (Fig. 1c and Supplementary Fig. 4e, f.). The anti-inflammatory response triggered by transferred sFc-stimulated hDC-SIGN⁺ bone-marrow-derived macrophages required the expression of the inhibitory Fc receptor (FcR) FcγRIIB, as FcγRIIB^{-/-} recipient mice were not protected from inflammation induced by K/BxN serum (Fig. 1d). Collectively, these results were consistent with the *in vivo* requirements for IVIG protection previously defined¹⁻⁴, and demonstrated that ligation of hDC-SIGN by sFc on bone-marrow-derived myeloid cells is sufficient to induce an anti-inflammatory cellular response.

DC-SIGN engagement has been reported to result in dendritic cell production of IL-10 (refs 7–8), making this anti-inflammatory cytokine an appealing candidate responsible for mediating IVIG anti-inflammatory activity. However, IL-10^{-/-} mice were protected from induced arthritis by IVIG similarly to wild-type controls (Supplementary Fig. 5). We next addressed other cytokines that could be responsible for this response. The T_H2 cytokine IL-4 has been shown to upregulate FcγRIIB surface expression on peripheral monocytes¹², and increase the threshold for activation by pathogenic immune complexes, consistent with the FcγRIIB requirement of IVIG^{1,13,14}. Therefore, sFc-treated DC-SIGN⁺ bone-marrow-derived macrophages were administered to naive wild-type mice or IL-4^{-/-} mice, and the recipient mice challenged with K/BxN serum. Wild-type recipients were protected from induced arthritis, whereas IL-4^{-/-} recipients were not (Fig. 2a).

These results led us to predict that IVIG anti-inflammatory activity would require IL-4 signalling. Indeed, mice deficient in IL-4 (IL-4^{-/-}; Fig. 2b), the IL-4 receptor (IL-4Rα^{-/-}; Fig. 2c), or the IL-4R signalling adaptor (Stat6^{-/-}; Fig. 2c), were not protected from K/BxN-induced inflammation by IVIG or sFc. Further, monocytes in the peripheral blood and bone marrow of wild-type mice, but not IL-4Rα^{-/-} mice, upregulated FcγRIIB after sFc administration (Supplementary Fig. 6). Next we examined whether exogenous T_H2 cytokines could also suppress autoantibody-induced inflammation. Mice were treated with cytokine immune complexes (ic)¹⁵ of T_H2 cytokines IL-4 (IL-4ic), and IL-13 (IL-13ic), or a non-T_H2 cytokine complex of IL-3 (IL-3ic), and challenged with K/BxN serum. Inflammation was significantly attenuated after single administration of IL-4ic or IL-13ic, but not after IL-3ic treatment (Fig. 2d). However, IL-4ic treatment did not attenuate

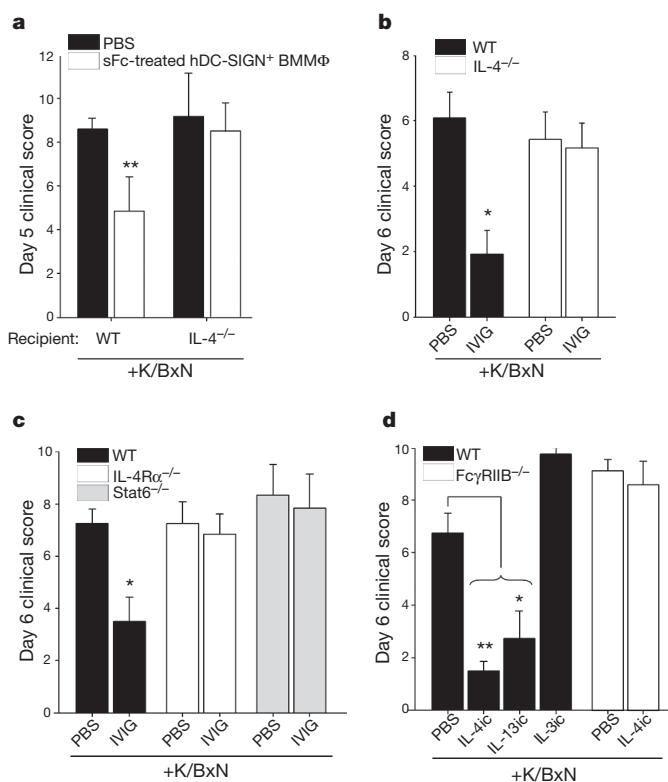


Figure 2 | IL-4 requirements of sFc anti-inflammatory activity. **a**, sFc-treated hDC-SIGN⁺ bone-marrow-derived macrophages (BMMΦ; white bars) or PBS was administered to K/BxN-treated wild-type or IL-4^{-/-} recipient mice. ***P* < 0.002 determined by Fisher LSD post-hoc test. **b**, Wild-type (black bars) and IL-4^{-/-} (white bars) mice were treated with K/BxN sera and IVIG. **P* < 0.01 determined by Mann–Whitney’s *U* test. **c**, Wild-type (black bars), IL-4Rα^{-/-} (white bars), or Stat6^{-/-} mice (grey bars) were given K/BxN sera and IVIG. **P* < 0.01 determined by Tukey’s post-hoc test. **d**, Wild-type (black bars) and FcγRIIB^{-/-} mice (white bars) were administered cytokine immune complexes (IL-4ic, IL-3ic, IL-13ic) and K/BxN sera. Means and standard deviations are plotted; **P* < 0.01, ***P* < 0.001 determined by Mann–Whitney’s *U* test.

inflammation in FcγRIIB^{-/-} mice, consistent with IL-4ic also requiring the FcγRIIB to suppress inflammation (Fig. 2d).

We next examined whether T_H2 cytokines were induced after IVIG or sFc administration. No changes in *Il4* messenger RNA levels were observed (Fig. 3a and Supplementary Fig. 7a–c), so we surveyed cytokines known to induce *Il4* expression, including IL-33 (refs 16–18), IL-25 (refs 18–20), and thymic stromal lymphopoietin (TSLP)^{18,21,22}. Interestingly, *Il33* mRNA was upregulated in wild-type mice after IVIG and sFc administration, but remained unchanged SIGN-R1^{-/-} mice. We then administered IL-33, IL-25, or TSLP to mice challenged with K/BxN serum. Exogenous IL-33 fully suppressed K/BxN arthritogenic activity and induced IL-4 production *in vivo*, whereas IL-25 promoted only modest protection, and TSLP provided no protection (Fig. 3b), all of which correlated with systemic IL-4 and IL-13 levels (Fig. 3b and Supplementary Fig. 7d, e). Further, exogenous IL-33 and IL-4ic were unable to ameliorate serum-induced arthritis in IL-4Rα^{-/-} mice (Fig. 3c and Supplementary Fig. 7e), indicating that IL-4Rα acts downstream of IL-33 in this pathway.

Our results support an anti-inflammatory cascade where DC-SIGN ligation by sFc promotes IL-33 production, IL-33 induces IL-4 expression, culminating in FcγRIIB upregulation on monocytes and macrophages (Supplementary Fig. 1). To confirm this, hDC-SIGN^{+/+}/SIGN-R1^{-/-} mice were treated with arthritogenic sera and sFc, in combination with a blocking antibody to the IL-33 receptor (anti-IL-33Rα). This intervention ablated the ability of sFc to protect

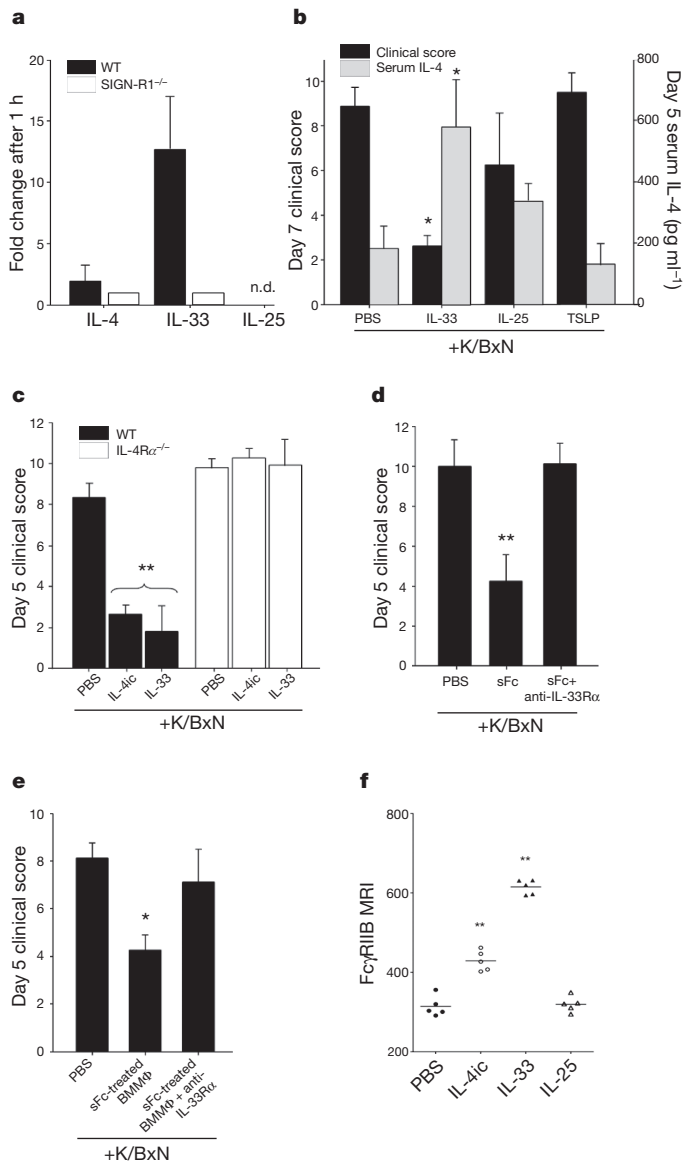


Figure 3 | IL-33 triggers IL-4 anti-inflammatory activity. **a**, Cytokine expression 1 h after IVIG administration in wild-type (black bars) or SIGN-R1^{-/-} mice (white bars) determined by quantitative polymerase chain reaction (qPCR). n.d., not detected. **b**, K/BxN-treated wild-type mice received PBS, IL-33, IL-25, or TSLP. **P* < 0.05 determined by Tukey's test. **c**, K/BxN-treated wild-type (black bars) or IL-4Rα^{-/-} mice (white bars) received PBS, IL-4ic, or IL-33. ***P* < 0.001 determined by Tukey's test. **d**, hDC-SIGN⁺/SIGN-R1^{-/-} mice received K/BxN sera, sFc and anti-IL-33Rα. ***P* < 0.001 determined by Fisher LSD test. **e**, sFc-treated hDC-SIGN⁺ bone-marrow-derived macrophages were administered to wild-type mice, K/BxN- and anti-IL-33Rα-treated wild-type mice. Means and standard deviations are plotted; **P* < 0.05 determined by Tukey's test. **f**, Individual mean fluorescence intensities (MFI) of bone marrow monocyte (CD11b⁺ Ly6G⁻) FcγRIIB surface expression 24 h after PBS, IL-4, IL-33, or IL-25 treatment by FACS. ***P* < 0.01 determined by Tukey's test.

hDC-SIGN⁺/SIGN-R1^{-/-} mice (Fig. 3d and Supplementary Fig. 7f). Protection of transferred sFc-treated hDC-SIGN⁺ bone-marrow-derived macrophages was also diminished by anti-IL-33Rα treatment (Fig. 3e). Further, administration of exogenous IL-4ic or IL-33 increased FcγRIIB surface expression on monocytes, whereas IL-25 had no effect (Fig. 3f and Supplementary Fig. 8). IL-4 treatment down-regulated FcγRIIB expression on B cells (Supplementary Fig. 8), consistent with the diverse effects of this cytokine on different leukocyte types.

IL-4 can be produced by T cells and several innate immune cell populations¹⁸, including basophils^{23,24}, mast cells^{25,26}, eosinophils²⁷ and progenitor cells¹⁹. IVIG activity is T-cell independent⁴, thus eliminating these cells as a source of sFc-induced IL-4. To determine whether basophils were involved in this response, we selectively depleted these cells *in vivo*²⁸ (Fig. 4a and Supplementary Figs 9–11). Arthritis could be induced in basophil-depleted hDC-SIGN⁺/SIGN-R1^{-/-} mice (Supplementary Fig. 9a) but the protective capacity of sFc and IVIG was lost (Fig. 4a and Supplementary Fig. 11), indicating that these cells have a pivotal role. We next administered sFc and K/BxN sera to IL-4-GFP reporter mice (4get²⁹), and found a twofold increase in GFP⁺ basophils (DX5⁺ FcεRI⁺ c-Kit⁻) in the circulation of protected sFc-treated mice, indicating that basophils produced IL-4 in response to sFc (Fig. 4b). To determine whether basophils were ultimately responsible for the anti-inflammatory activity induced by sFc through IL-33, we transferred PBS- or IL-33-treated basophils to K/BxN-treated recipient mice (Fig. 4c, d and Supplementary Fig. 12a–c). IL-33-treated basophils derived from wild-type or FcγRIIB^{-/-} mice were equally effective at suppressing arthritic inflammation, reducing serum IL-6 levels, and curbing leukocyte infiltration to arthritic paws (Fig. 4c, d and Supplementary Fig. 12d, e). These results confirm the anti-inflammatory potential of these cells, and support a model whereby IL-4 produced by basophils increases FcγRIIB expression on inflammatory macrophages (Supplementary Fig. 1).

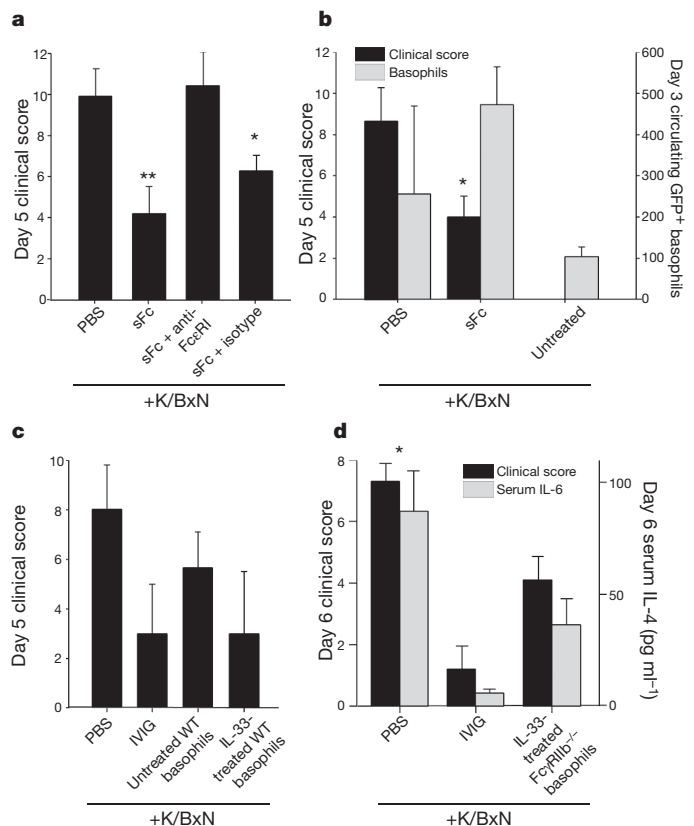


Figure 4 | Anti-inflammatory activity mediated by basophils. **a**, hDC-SIGN⁺/SIGN-R1^{-/-} mice were administered K/BxN sera, sFc and anti-FcεRI or an isotype control. ***P* < 0.001, **P* < 0.05 determined by Fisher LSD test. **b**, 4get mice were administered K/BxN sera and sFc. Circulating IL-4⁺ basophils (grey bars, DX5⁺ FcεRI⁺ GFP⁺) and clinical scores (black bars) are plotted. **P* < 0.05 determined by Tukey's test. **c**, PBS or IL-33-treated basophils (DX5⁺ FcεRI⁺ c-Kit⁻) were administered to K/BxN-treated wild-type mice. Control mice received PBS or IVIG. **d**, Basophils from IL-33-treated FcγRIIB^{-/-} mice were administered to K/BxN-treated wild-type recipients. Clinical scores (black) and serum IL-6 levels (grey) are plotted. Means and standard deviations are plotted; **P* < 0.05 determined by Mann-Whitney's U test.

Analysing the anti-inflammatory activity of IVIG led to identification of an endogenous, innate pathway in which sialylated IgG, a minor component of serum IgG antibodies, binds DC-SIGN, promoting production of IL-33, which expands IL-4⁺ basophils. These cytokines are capable of suppressing autoantibody-mediated inflammation by modulating FcγRIIB expression on effector cells (Supplementary Fig. 1). IL-4 and IL-33 have pleiotropic activities, and mediate T_H2 responses to helminth parasites and allergens¹⁸, as well as enhance inflammatory arthritis^{30,31}, in addition to their activities reported here. Cytokine concentration, cellular environment, and differential responses of individual cell types are likely to explain these distinct effector functions.

Various stimuli have been reported to modulate the level of IgG Fc sialylation, and could regulate this intrinsic pathway. Antigenic stimulation results in the production of pro-inflammatory, antigen-specific, asialylated IgG antibodies². Pathogenic autoantibodies, such as those produced during rheumatoid arthritis that recognize citrullinated peptides, similarly show reduced sialic acid as compared to other serum antibodies³². Conversely, increases in sialylated IgG antibodies occur during pregnancy³³, which could contribute to the remission in arthritis seen in pregnant women. Therefore, affecting the sialylation of IgG antibodies could provide an intrinsic mechanism for regulating T_H2 cytokine production by innate myeloid cells in a DC-SIGN-dependent manner, provide a means for maintaining homeostasis, and is an attractive therapeutic approach to suppressing inflammation in autoimmune diseases.

METHODS SUMMARY

hDC-SIGN BAC transgenic mice were created using the BAC clone CTD2102F19 (Invitrogen) containing the human *DC-SIGN* gene. The BAC was linearized by the NotI restriction endonuclease. The human *DC-SIGN* gene fragment was purified and injected into one-day-old C57BL/6 embryos via pronuclear microinjection. The embryos were then implanted into ICR surrogate females and the resulting progeny were screened by PCR for the presence of the human *DC-SIGN* transgene. hDC-SIGN⁺ mice were crossed to SIGN-R1^{-/-} mice to generate hDC-SIGN⁺/SIGN-R1^{-/-} lines.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 1 October 2010; accepted 20 April 2011.

Published online 19 June; corrected 7 July 2011 (see full-text HTML version for details).

1. Nimmerjahn, F. & Ravetch, J. V. Anti-inflammatory actions of intravenous immunoglobulin. *Annu. Rev. Immunol.* **26**, 513–533 (2008).
2. Kaneko, Y., Nimmerjahn, F. & Ravetch, J. V. Anti-inflammatory activity of immunoglobulin G resulting from Fc sialylation. *Science* **313**, 670–673 (2006).
3. Anthony, R. M. *et al.* Recapitulation of IVIG anti-inflammatory activity with a recombinant IgG Fc. *Science* **320**, 373–376 (2008).
4. Anthony, R. M., Wermeling, F., Karlsson, M. C. & Ravetch, J. V. Identification of a receptor required for the anti-inflammatory activity of IVIG. *Proc. Natl Acad. Sci. USA* **105**, 19571–19578 (2008).
5. Granelli-Piperno, A. *et al.* Dendritic cell-specific intercellular adhesion molecule 3-grabbing nonintegrin/CD209 is abundant on macrophages in the normal human lymph node and is not required for dendritic cell stimulation of the mixed leukocyte reaction. *J. Immunol.* **175**, 4265–4273 (2005).
6. Soilleux, E. J. *et al.* Constitutive and induced expression of DC-SIGN on dendritic cell and macrophage subpopulations *in situ* and *in vitro*. *J. Leukoc. Biol.* **71**, 445–457 (2002).
7. Geijtenbeek, T. B. & Gringhuis, S. I. Signalling through C-type lectin receptors: shaping immune responses. *Nature Rev. Immunol.* **9**, 465–479 (2009).
8. Gringhuis, S. I. *et al.* C-type lectin DC-SIGN modulates Toll-like receptor signaling via Raf-1 kinase-dependent acetylation of transcription factor NF-κB. *Immunity* **26**, 605–616 (2007).
9. Hodges, A. *et al.* Activation of the lectin DC-SIGN induces an immature dendritic cell phenotype triggering Rho-GTPase activity required for HIV-1 replication. *Nature Immunol.* **8**, 569–577 (2007).
10. Korganow, A. S. *et al.* From systemic T cell self-reactivity to organ-specific autoimmune disease via immunoglobulins. *Immunity* **10**, 451–461 (1999).

11. Schaefer, M. *et al.* Decreased pathology and prolonged survival of human DC-SIGN transgenic mice during mycobacterial infection. *J. Immunol.* **180**, 6836–6845 (2008).
12. Pricop, L. *et al.* Differential modulation of stimulatory and inhibitory Fcγ receptors on human monocytes by Th1 and Th2 cytokines. *J. Immunol.* **166**, 531–537 (2001).
13. Bruhns, P., Samuelsson, A., Pollard, J. W. & Ravetch, J. V. Colony-stimulating factor-1-dependent macrophages are responsible for IVIG protection in antibody-induced autoimmune disease. *Immunity* **18**, 573–581 (2003).
14. Samuelsson, A., Towers, T. L. & Ravetch, J. V. Anti-inflammatory activity of IVIG mediated through the inhibitory Fc receptor. *Science* **291**, 484–486 (2001).
15. Finkelman, F. D. *et al.* Anti-cytokine antibodies as carrier proteins. Prolongation of *in vivo* effects of exogenous cytokines by injection of cytokine-anti-cytokine antibody complexes. *J. Immunol.* **151**, 1235–1244 (1993).
16. Schmitz, J. *et al.* IL-33, an interleukin-1-like cytokine that signals via the IL-1 receptor-related protein ST2 and induces T helper type 2-associated cytokines. *Immunity* **23**, 479–490 (2005).
17. Neill, D. R. *et al.* Nuocytes represent a new innate effector leukocyte that mediates type-2 immunity. *Nature* **464**, 1367–1370 (2010).
18. Paul, W. E. & Zhu, J. How are T_H2-type immune responses initiated and amplified? *Nature Rev. Immunol.* **10**, 225–235 (2010).
19. Saenz, S. A. *et al.* IL25 elicits a multipotent progenitor cell population that promotes T_H2 cytokine responses. *Nature* **464**, 1362–1366 (2010).
20. Fort, M. M. *et al.* IL-25 induces IL-4, IL-5, and IL-13 and Th2-associated pathologies *in vivo*. *Immunity* **15**, 985–995 (2001).
21. Soumelis, V. *et al.* Human epithelial cells trigger dendritic cell mediated allergic inflammation by producing TSLP. *Nature Immunol.* **3**, 673–680 (2002).
22. Wang, Y. H. *et al.* IL-25 augments type 2 immune responses by enhancing the expansion and functions of TSLP-DC-activated Th2 memory cells. *J. Exp. Med.* **204**, 1837–1847 (2007).
23. Min, B. *et al.* Basophils produce IL-4 and accumulate in tissues after infection with a Th2-inducing parasite. *J. Exp. Med.* **200**, 507–517 (2004).
24. Seder, R. A. *et al.* Mouse splenic and bone marrow cell populations that express high-affinity Fc epsilon receptors and produce interleukin 4 are highly enriched in basophils. *Proc. Natl Acad. Sci. USA* **88**, 2835–2839 (1991).
25. Seder, R. A. *et al.* Production of interleukin-4 and other cytokines following stimulation of mast cell lines and *in vivo* mast cells/basophils. *Int. Arch. Allergy Appl. Immunol.* **94**, 137–140 (1991).
26. Wang, M., Saxon, A. & Diaz-Sanchez, D. Early IL-4 production driving Th2 differentiation in a human *in vivo* allergic model is mast cell derived. *Clin. Immunol.* **90**, 47–54 (1999).
27. Shinkai, K., Mohrs, M. & Locksley, R. M. Helper T cells regulate type-2 innate immunity *in vivo*. *Nature* **420**, 825–829 (2002).
28. Sokol, C. L., Barton, G. M., Farr, A. G. & Medzhitov, R. A mechanism for the initiation of allergen-induced T helper type 2 responses. *Nature Immunol.* **9**, 310–318 (2008).
29. Mohrs, M., Shinkai, K., Mohrs, K. & Locksley, R. M. Analysis of type 2 immunity *in vivo* with a bicistronic IL-4 reporter. *Immunity* **15**, 303–311 (2001).
30. Ohmura, K. *et al.* Interleukin-4 can be a key positive regulator of inflammatory arthritis. *Arthritis Rheum.* **52**, 1866–1875 (2005).
31. Xu, D. *et al.* IL-33 exacerbates autoantibody-induced arthritis. *J. Immunol.* **184**, 2620–2626 (2010).
32. Scherer, H. U. *et al.* Glycan profiling of anti-citrullinated protein antibodies isolated from human serum and synovial fluid. *Arthritis Rheum.* **62**, 1620–1629 (2010).
33. van de Geijn, F. E. *et al.* Immunoglobulin G galactosylation and sialylation are associated with pregnancy-induced improvement of rheumatoid arthritis and the postpartum flare: results from a large prospective cohort study. *Arthritis Res. Ther.* **11**, R193 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The authors thank P. Smith, M. Kibe, J. Brown and K. Velinon for technical support, K. L. Jeffrey, C. Cheong, R. Steinman and J. J. Lee for discussions, M. Pack for providing human spleen sections, A. McKenzie for providing SIGN-R1^{-/-} mice, T. Sparwasser for providing CD11c-hDC-SIGN mice, C. G. Park for providing hDC-SIGN and hDC-SIGN-R expressing cell lines, and H. Watarai for providing anti-IL-25R antibodies. R.M.A. is an Irvington Institute fellow of the Cancer Research Institute. F.W. is supported by the Wenner-Gren Foundations, Sweden. This work was performed with support from Virdante Pharmaceuticals and NIH grants to J.V.R.

Author Contributions R.M.A., T.K., F.W. and J.V.R. designed the experiments and interpreted the results. R.M.A., T.K. and F.W. performed the experiments, and R.M.A. and J.V.R. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.V.R. (ravetch@rockefeller.edu).

METHODS

Mice. Eight-to-twelve-week old, sex and age matched mice were used for all experiments in compliance with federal laws, institutional guidelines and have been approved by the Rockefeller University. Wild-type C57BL/6, wild-type BALB/c, NOD, IL-4^{-/-}, IL-4Rα^{-/-}, Stat6^{-/-}, IL-10^{-/-}, 4get (IL-4-GFP reporter) mice were purchased from Jackson Laboratories, and maintained at the Rockefeller University animal facility. FcγRIIB^{-/-} mice³⁴ were generated previously in the laboratory. SIGN-R1^{-/-} mice³⁵ were provided by A. McKenzie. CD11c-hDCSIGN⁺ mice¹¹ were provided by T. Sparwasser. KRN T-cell receptor transgenic mice on a C57BL/6 background (K/B) were gifts from D. Mathis and C. Benoist and were bred to NOD mice to generate K/BxN mice¹⁰. K/BxN serum was prepared as described previously¹³. Briefly, serum was separated from blood collected from the K/BxN mice (6–12-weeks old). Several weeks of serum collection were pooled together and frozen in aliquots to be used in the experiments described here. One intravenous injection of 200 μl K/BxN serum was used to induce arthritis. Severity of arthritis was scored by clinical examination by adding the index of all four paws, where 0 is unaffected, 1 is swelling of one joint, 2 is swelling of more than one joint, and 3 is severe swelling of the entire paw. All experiments shown yielded similar results at least 3 times with treatment groups of 4–5 mice, and means and standard deviations are plotted in bar graphs.

hDC-SIGN BAC transgenic mice were created using the BAC clone CTD2102F19 (Invitrogen) containing the human *DC-SIGN* gene. The BAC was linearized by the NotI restriction endonuclease. The human *DC-SIGN* gene fragment was purified and injected into one-day-old C57BL/6 embryos via pronuclear microinjection. The embryos were then implanted into ICR surrogate females and the resulting progeny were screened by PCR for the presence of the human *DC-SIGN* transgene. hDC-SIGN⁺ mice were crossed to SIGN-R1^{-/-} mice to generate hDC-SIGN⁺/SIGN-R1^{-/-} lines.

Reagents and treatments. IVIG (Octagam, Octapharma) or IVIG-derived Fcs was enriched for terminal sialic acid using SNA-agarose² (Vector Laboratories) or hypersialated *in vitro* as previously described³ to generate sFc. AsialoFc was generated by treating Fcs with neuraminidase (NEB) as per the manufacturer's directions. Sialic acid content was verified by lectin blotting with SNA-biotin (Vector Laboratories). IVIG and IVIG derivations were administered intravenously (i.v.) at 1 g kg⁻¹, SNA-enriched IVIG at 0.1 g kg⁻¹, and sFc at 0.03 g kg⁻¹ one hour before K/BxN sera administration.

Mice receiving cytokine:immune complexes (ic) with prolonged half-life were treated with a single i.v. injection 2.5 μg of cytokine (IL-3, IL-4, IL-13; Peprotech) and 12.5 μg of neutralizing antibody at day 0. Neutralizing antibodies used were anti-IL-3 (MP2-8F8, Biolegend), anti-IL-4 (11B11, BD Biosciences), and anti-IL-13 (eBio1316H, eBioscience). Other cytokine treatments included intraperitoneal (i.p.) administration of 400 ng (or 800 ng) of IL-25 (R&D), 400 ng of IL-33 (R&D), or 1 μg of TSLP (R&D) on days 0, 1, 2, and 3. Basophils were depleted as described²⁸ by daily i.p. injection with 10 μg of anti-FcεRI (MAR-1, eBioscience) or hamster IgG isotype control (eBioscience) on days 0–5. Alternatively, mice received a single i.v. injection of 30 μg anti-CD200RL3 (ref. 36) (Ba103, Hycult Biotech) or rat IgG isotype control (BD Biosciences). IL-33Rα was blocked by i.v. injection of 80 μg of anti-IL-33Rα (DT8, MD Biosciences) or rat IgG1 isotype control (BD Biosciences) on day 0. hDC-SIGN was blocked *in vivo* by administration of 125 μg E9E A8 (ref. 37) or isotype control mouse IgG2a (BioLegend).

IL-6 was measured in serum by ELISA as suggested by the manufacturer (BioLegend). Serum IL-4 and IL-13 was measured using an *in vivo* cytokine capture assay as described³⁸. Briefly, 10 μg biotinylated anti-IL-4 antibody (clone BVD4-1D11, BioLegend) or biotinylated anti-IL-13 (eBio1316HA, eBioscience) was injected i.v. into treated mice, and sera were collected 24 h later. Cytokine levels were quantified by ELISA assay using anti-IL-4 (BVD6-24G1, BioLegend) or anti-IL-13 (eBio13A, eBioscience) as capture antibodies.

Splenic RNA was purified using RNeasy Mini Kits (Qiagen) and reverse-transcribed using Verso cDNA synthesis kit (Thermo Scientific). Quantitative PCR (qPCR) was conducted in 7300 Real-time PCR System (Life Technologies) with primer-probe sets for mouse IL-4, IL-13, IL-33, IL-25, or rRNA (Life Technologies), and gene expression levels were determined by normalization to rRNA levels.

Saturation binding experiments were performed as previously described⁴, comparing CHO and CHO-hDC-SIGN cells or Hep-CD81 and Hep-hDC-SIGN-R cells.

Flow cytometry. Single cell suspensions were prepared from peripheral blood, spleen, bone marrow, or paws from mice. After red blood cell lysis, cells were stained with the indicated monoclonal antibodies, and subjected to analysis using a FACSCalibur or LSR-II cytometer (BD Biosciences). Human leukocytes were

obtained from peripheral blood samples (New York Blood Center) after density gradient centrifugation (Ficoll-Paque, GE Healthcare). Antibodies used for murine cell staining were as follows: anti-CD19 (1D3), anti-B220 (RA3-6B2), anti-CD3ε (145-2C11), anti-CD11b (M1/70), anti-Ly6G (1A8), anti-CD11c (HL3), anti-I-A^b (AF6-120.1), anti-CD49b (DX5 and HMA2), anti-c-Kit (2B8), anti-CD45.2 (104) from BD Biosciences, anti-NKp46 (29A1.4), anti-SIGN-R1 (22D1), anti-CD123 (5B11) from eBioscience, anti-hDC-SIGN (9E9A8), anti-FcεRI (MAR-1) from Biolegend, anti-FcγRIIB (K9.361), anti-hDC-SIGNR (120604) from R&D systems. Antibodies used for human cell staining were: anti-CD14 (M5E2), anti-CD16 (B73.1), anti-CD3 (UCHT1), anti-CD56 (B159), anti-CD19 (SJ25C1), anti-CD11c (B-Ly6), anti-HLA-DR (L243 (G46-6)), anti-hDC-SIGN (AZND1) from BD Biosciences, and anti-hDC-SIGN (9E9A8, Biolegend). AccuCheck Counting Beads (Invitrogen) were used to quantify cells.

Bone-marrow macrophage and dendritic cell cultures and transfers. Bone-marrow-derived macrophages were cultured as described previously³⁹. Briefly, marrow was recovered from tibias and femurs of mice, and seeded in non-tissue culture treated 10-cm plates with DMEM supplemented with 10% fetal bovine serum, 2% penicillin/streptomycin (Invitrogen), 1% glutamine 200 mM (Invitrogen), 0.1% β-mercaptoethanol, IL-3 (5 ng ml⁻¹, Peprotech) and M-CSF (5 ng ml⁻¹, Peprotech) overnight at 37 °C, 5% CO₂. The next day, non-adherent cells were recovered and plated in 10-cm non-tissue culture treated plates in supplemented DMEM with cytokines, and cultured for 5–7 days. Once the cultured cells were mature macrophages (>90% CD11b⁺ F4/80⁺ by FACS), the cells were detached and 2 × 10⁶ macrophages were plated per well in 6-well plates, and allowed to attach overnight. The next day, the cells were pulsed with IVIG (15 mg ml⁻¹), BSA (15 mg ml⁻¹, Sigma), sFc (0.5 mg ml⁻¹), or asialoFc (0.5 mg ml⁻¹) for 30 min at 37 °C. The cells were recovered, washed thoroughly in cold PBS, and 1 × 10⁶ macrophages were administered i.v. into naive recipients. One hour later, the recipient mice were treated with K/BxN sera. Dendritic cells were cultured from mouse tibia and femur bone marrow cells as described⁴⁰. Briefly, 1 × 10⁶ cells ml⁻¹ were plated in 24-well plates with DMEM supplemented with 10% FBS and 10 ng ml⁻¹ mouse granulocyte macrophage colony stimulating factor (GM-CSF, Peprotech). On day 6, loosely adherent cells were collected by gentle pipetting, and were subjected to flow cytometric analysis or bone marrow cell transfer experiments as described above.

Histology. Human lymph node samples were from ISL Bio, and M. Pack provided human spleen samples. Spleens or lymph nodes embedded in O.C.T. compound (Sakura Finetek) were fixed in ice-cold acetone for 10 min, and stained with anti-SIGN-R1, anti-hDC-SIGN or anti-hDC-SIGN-R for 1 h at 25 °C in combination with antibodies for macrophages or B cells. Antibodies used included anti-F4/80 (BM8, Invitrogen) for mouse red pulp macrophages, anti-B220 for mouse B cells, anti-hCD20 (2H7, Biolegend) for human B cells, and anti-CD68 (Y1/82A, Biolegend) for human macrophages. Sections were visualized by wide-field fluorescence microscope (Zeiss).

Basophil adoptive transfers. Basophils were expanded by administering IL-3ic⁴¹, as described above, to wild-type or FcγRIIB^{-/-} mice. Five days later, IL-3ic-treated mice were administered PBS or IL-33 (400 ng) i.p. The next day, basophils (DX5⁺ FcRI⁺ cKit⁻) were sorted using a FACSaria II (BD Biosciences). Sorted basophils were washed in cold PBS, and 0.7 × 10⁶ basophils were administered to naive recipient mice subsequently administered K/BxN sera. A Wright–Giemsa stain of sorted, cytopun basophils was performed as suggested by the manufacturer (Sigma).

34. Takai, T., Ono, M., Hikida, M., Ohmori, H. & Ravetch, J. V. Augmented humoral and anaphylactic responses in FcγRII-deficient mice. *Nature* **379**, 346–349 (1996).
35. Lanoue, A. *et al.* SIGN-R1 contributes to protection against lethal pneumococcal infection in mice. *J. Exp. Med.* **200**, 1383–1393 (2004).
36. Obata, K. *et al.* Basophils are essential initiators of a novel type of chronic allergic inflammation. *Blood* **110**, 913–920 (2007).
37. Cheong, C. *et al.* New monoclonal anti-mouse DC-SIGN antibodies reactive with acetone-fixed cells. *J. Immunol. Methods* **360**, 66–75 (2010).
38. Finkelman, F., Morris, S., Orekhova, T. & Sehly, D. The *in vivo* cytokine capture assay for measurement of cytokine production in the mouse. *Curr. Protoc. Immunol.* **6.28.1–6.28.10** (2003).
39. Jeffrey, K. L. *et al.* Positive regulation of immune cell function and inflammatory responses by phosphatase PAC-1. *Nature Immunol.* **7**, 274–283 (2006).
40. Inaba, K. *et al.* Generation of large numbers of dendritic cells from mouse bone marrow cultures supplemented with granulocyte/macrophage colony-stimulating factor. *J. Exp. Med.* **176**, 1693–1702 (1992).
41. Ohmori, K. *et al.* IL-3 induces basophil expansion *in vivo* by directing granulocyte-monocyte progenitors to differentiate into basophil lineage-restricted progenitors in the bone marrow and by increasing the number of basophil/mast cell progenitors in the spleen. *J. Immunol.* **182**, 2835–2841 (2009).

XUTs are a class of Xrn1-sensitive antisense regulatory non-coding RNA in yeast

E. L. van Dijk^{1*}, C. L. Chen^{1*}, Y. d'Aubenton-Carafa¹, S. Gourvennec², M. Kwapisz², V. Roche², C. Bertrand², M. Silvain¹, P. Legoux-Né³, S. Loeillet⁴, A. Nicolas⁴, C. Thermes¹ & A. Morillon^{1,2}

Non-coding (nc)RNAs are key players in numerous biological processes such as gene regulation, chromatin domain formation and genome stability^{1,2}. Large ncRNAs interact with histone modifiers^{3–5} and are involved in cancer development⁶, X-chromosome inactivation⁷ and autosomal gene imprinting⁸. However, despite recent evidence showing that pervasive transcription is more widespread than previously thought⁹, only a few examples mediating gene regulation in eukaryotes have been described¹⁰. In *Saccharomyces cerevisiae*, the bona-fide regulatory ncRNAs are destabilized by the Xrn1 5'–3' RNA exonuclease^{11,12} (also known as Kem1), but the genome-wide characterization of the entire regulatory ncRNA family remains elusive. Here, using strand-specific RNA sequencing (RNA-seq), we identify a novel class of 1,658 Xrn1-sensitive unstable transcripts (XUTs) in which 66% are antisense to open reading frames. These transcripts are polyadenylated and RNA polymerase II (RNAPII)-dependent. The majority of XUTs strongly accumulate in lithium-containing media, indicating that they might have a role in adaptive responses to changes in growth conditions. Notably, RNAPII chromatin immunoprecipitation followed by DNA sequencing (ChIP-seq) analysis of Xrn1-deficient strains revealed a significant decrease of RNAPII occupancy over 273 genes with antisense XUTs. These genes show an unusual bias for H3K4me3 marks and require the Set1 histone H3 lysine 4 methyl-transferase for silencing. Furthermore, abolishing H3K4me3 triggers the silencing of other genes with antisense XUTs, supporting a model in which H3K4me3 antagonizes antisense ncRNA repressive activity. Our results demonstrate that antisense ncRNA-mediated regulation is a general regulatory pathway for gene expression in *S. cerevisiae*.

Antisense ncRNAs in yeast control Ty1 mobility^{12,13} and *PHO84* transcription through histone modifications^{11,12} in an RNA-interference-independent manner¹⁴ (Supplementary Fig. 1). The regulatory ncRNAs are distinct from the canonical cryptic unstable transcripts (CUTs) of shorter size, destabilized by the nuclear exosome¹⁵. In contrast, the regulatory antisense ncRNAs are mainly degraded by the cytoplasmic 5'–3' Xrn1 exonuclease¹⁶. These features suggest that yeast antisense regulatory ncRNAs belong to a separate class of unstable transcripts. Here, we address their systematic identification by characterizing the Xrn1-dependent cryptic transcriptome.

We analysed the whole transcriptome in wild-type and *xrn1Δ* strains (Fig. 1a). Because Xrn1 mainly acts downstream of RNA deadenylation¹⁶, we analysed total RNA depleted for ribosomal RNAs (ribo[–]) and subsequently compared with fractions enriched for polyadenylated RNAs (polyA⁺) (Supplementary Fig. 2). Ultra-deep sequencing produced 84 and 134 millions of unique reads for ribo[–] preparations from wild-type and *xrn1Δ* strains, respectively (Supplementary Table 1), reaching unprecedented coverage of the yeast cryptic transcriptome^{17–19}. Read densities were normalized with transfer RNAs and small nucleolar RNAs insensitive to Xrn1 activity^{20,21} (Methods and

Supplementary Fig. 3). Similar results were obtained with biological duplicates and with ABI-SOLiD or Illumina sequencing technologies (Supplementary Fig. 4a, b). Transcripts were identified along read density profiles and compared to annotated open reading frames (ORFs) and stable uncharacterized transcripts (SUTs)²² (Methods and Supplementary Fig. 5). The Xrn1-deficient strain showed an average 4.4- to 3.9-fold increase of ORF and dubious ORF transcripts, respectively (Fig. 1b and Supplementary Fig. 6a, b), further establishing Xrn1 as the major exonuclease responsible for messenger RNA turnover¹⁶. The data revealed 932 previously uncharacterized ncRNAs (New) presenting a 10.8-fold average increase in the *xrn1Δ* strain (Fig. 1b, c and Supplementary Fig. 6c). Additionally, 75% (543) of the SUTs described in rich media²² showed an unexpected 9.6-fold increase in the *xrn1Δ* strain, indicating that Xrn1 is also active on these ncRNAs (Fig. 1c and Supplementary Fig. 6d). In contrast to the majority of CUTs, 20% (183) of them were highly sensitive to Xrn1 (7.9-fold increase, Fig. 1c and Supplementary Fig. 6e), showing that a subclass of CUTs escapes from exosome-dependent nuclear degradation, as previously suggested²³. Thus, we describe 1,658 ncRNAs that are Xrn1-sensitive unstable transcripts (XUTs), of which 66% are antisense to ORFs (Fig. 1b, c). The start sites of three XUTs upstream of the *SUC2* locus (*SUC2uxut*) (Fig. 1d) were identified by 5' rapid amplification of cDNA ends (5'RACE) experiments (Supplementary Fig. 7). Northern blot analyses (Fig. 1e) showed that the levels of *SUC2uxut1* and *SUC2uxut3* are independent of Trf4, subunit of the TRAMP complex, required for CUT degradation²⁴. This confirms that XUTs encompass a distinct class of cryptic transcripts. Further analyses showed that XUTs are polyadenylated and synthesized by RNAPII (Methods and Supplementary Fig. 8a–d).

We asked whether changes in growth conditions would activate XUTs in wild-type cells. Interestingly, lithium toxicity is due to inactivation of Xrn1, Rat1 and RNase MRP²⁵ and might be associated with XUT overexpression. The kinetics of appearance of selected XUTs in lithium-containing media (Fig. 2a and Supplementary Fig. 9) showed that the Ty1-ncRNA (*RTL*) appears before 30 min, similarly to *SUC2uxut1* and *TPO1axut* (transcriptome details in Supplementary Fig. 10a). After washing and further growth, cells present a dramatic decrease of XUTs levels, indicating a highly dynamic XUT accumulation. RNA-seq analyses after 80 min in lithium media showed XUTs overexpression (Fig. 2b) and 56% of them have similar RNA levels as the *xrn1Δ* strain (Fig. 2c). To note, some XUTs present a milder sensitivity to lithium as exemplified by *SUC2uxut2* and *SUC2uxut3* (Fig. 2a, d). These results indicate that a majority of XUTs rapidly and reversibly accumulate after lithium addition, indicating a role in the early response to this growth media change.

We tested whether antisense ncRNA accumulation controls sense gene transcription as for Ty1 (ref. 12). ChIP-seq analyses identified 996 ORFs with reduced RNAPII occupancy in the *xrn1Δ* strain (Fig. 3a, $P < 0.01$, Methods); among them, 273 (27%) accumulate antisense

¹Centre de Génétique Moléculaire (CNRS UPR 3404), avenue de la Terrasse, 91198 Gif sur Yvette, France. ²ncRNA, epigenetic and genome fluidity, Institut Curie, Centre de recherche, CNRS UMR3244, Université Pierre et Marie Curie, 26 rue d'Ulm, 75248 Paris Cedex 05, France. ³NGS Platform, Institut Curie, 26 rue d'Ulm, 75248 Paris Cedex 05, France. ⁴Recombination and Genome instability, Institut Curie, Centre de recherche, CNRS UMR3244, Université Pierre et Marie Curie, 26 rue d'Ulm, 75248 Paris Cedex 05, France.

*These authors contributed equally to this work.

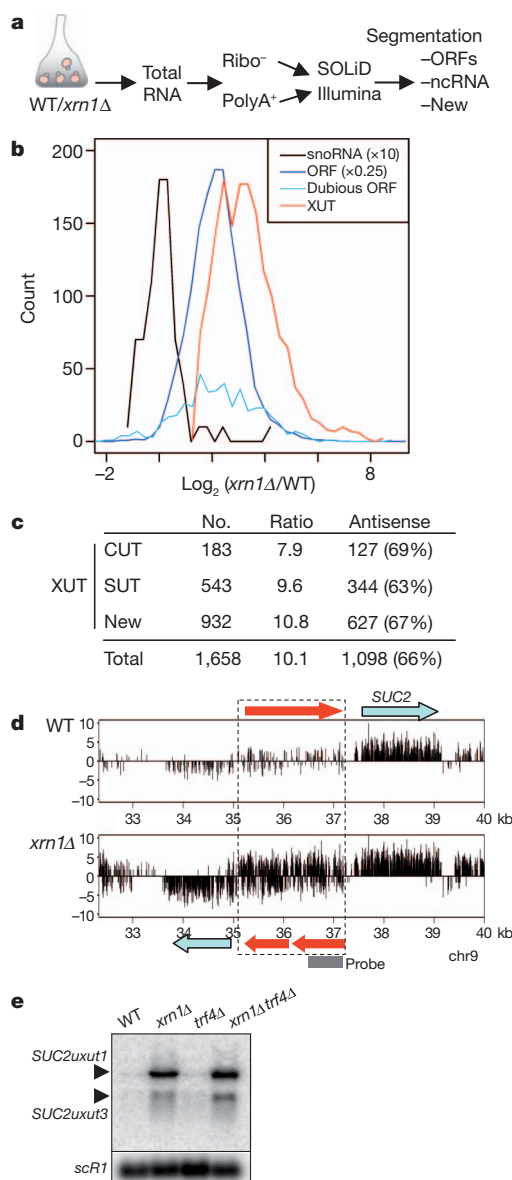


Figure 1 | Cytoplasmic 5'-3' RNA decay controls a class of cryptic ncRNA.

a, RNA-seq analyses using wild-type (WT) and *xrn1Δ* strains. Cells were grown in rich media and cDNA sequenced with SOLiD or Illumina technologies (Methods). Segmentation defines transcriptional units (ORFs, ncRNAs, unidentified transcripts (New)). **b**, XUTs define a novel class of ncRNAs. Tag density ratio distributions of *xrn1Δ*/WT for snoRNAs, ORFs and XUTs. **c**, Numbers of XUTs including new ncRNAs (New), most SUTs and few CUTs; average ratios of *xrn1Δ*/WT, numbers and percentage of XUTs antisense to ORFs. **d**, Transcriptome data for *SUC2* locus. Three XUTs (red) were identified within *SUC2* promoter in both orientations; bars, log₂ read numbers (upward and downward, Watson and Crick strands, respectively); grey bar, probe used in **e**. **e**, Northern blot for indicated strains of sense *SUC2uxut1* (2.3 kb) and antisense *SUC2uxut3* (1 kb), using a double-stranded probe (**d**) and control *scR1*.

XUTs (Fig. 3a, class 1, Supplementary Table 2). This proportion is larger than for the ORFs that do not present significantly reduced RNAPII levels ($P < 10^{-15}$, chi-squared test; only 15% associate with antisense XUTs), indicating that sense transcriptional silencing correlates with antisense RNA accumulation. Although RNA stabilization in *xrn1Δ* strain could mask potential transcriptional repression, we identified 72 downregulated genes with antisense XUTs (Supplementary Fig. 12) among which 65 showed reduced RNAPII occupancy (Supplementary Fig. 13 and Supplementary Table 3). We conclude that, even though RNA-seq is more suitable to characterize XUT transcripts, ChIP-seq analyses identify extensively downregulated genes

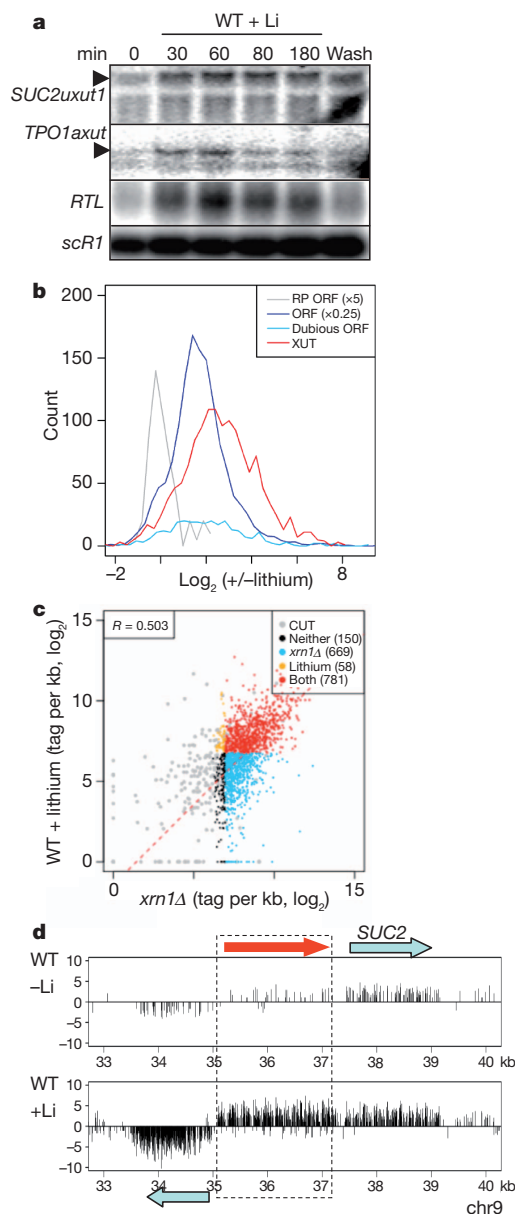


Figure 2 | XUT ncRNAs accumulate in wild-type cells grown in presence of lithium.

a, XUT levels rapidly increase upon lithium treatment. Northern blot of total RNA from wild-type cells grown without and with 100 mM lithium at different times (Methods); strand-specific probes for *SUC2uxut1* and *SUC2uxut3* as in Fig. 1e, *TPO1axut* and *TY1* XUTs (*RTL*). **b**, XUTs are overexpressed in lithium media. Same as in Fig. 1b with wild type with or without lithium. Ribosomal protein genes were used as reference (grey, Methods). **c**, Tag density of XUTs and CUTs in *xrn1Δ* (abscissa) and WT + lithium (ordinate); 56% of XUTs are similarly expressed in *xrn1Δ* and WT grown in lithium (80 min). XUT expression compared to CUT expression (grey). In the subset of XUTs expressed at higher level than 75% of CUTs, 669 are found in *xrn1Δ* only (blue), 58 in lithium only (yellow) and 781 in both conditions (red). **d**, Same as in Fig. 1d with wild-type cells with or without lithium.

like *TIR1* and *TPO1* (Fig. 3b and Supplementary Fig. 10c, transcriptomes in Fig. 4b and Supplementary Fig. 10a). In these cases, RNAPII is almost undetectable within the XUTs' loci in the *xrn1Δ* strain, indicating that either antisense transcription is also silenced or XUT transcription is too low for detection. To establish that XUT accumulation and gene silencing are direct consequences of *XRN1* deletion, we performed *xrn1* conditional inactivation (*xrn1ts²⁶*, Methods). Not only do 70% of XUTs share identical expression in *xrn1Δ* and *xrn1ts* strains (Supplementary Fig. 14), but both strains also show similar RNAPII occupancies on a selection of class 1 genes (Figure 3c). This suggests

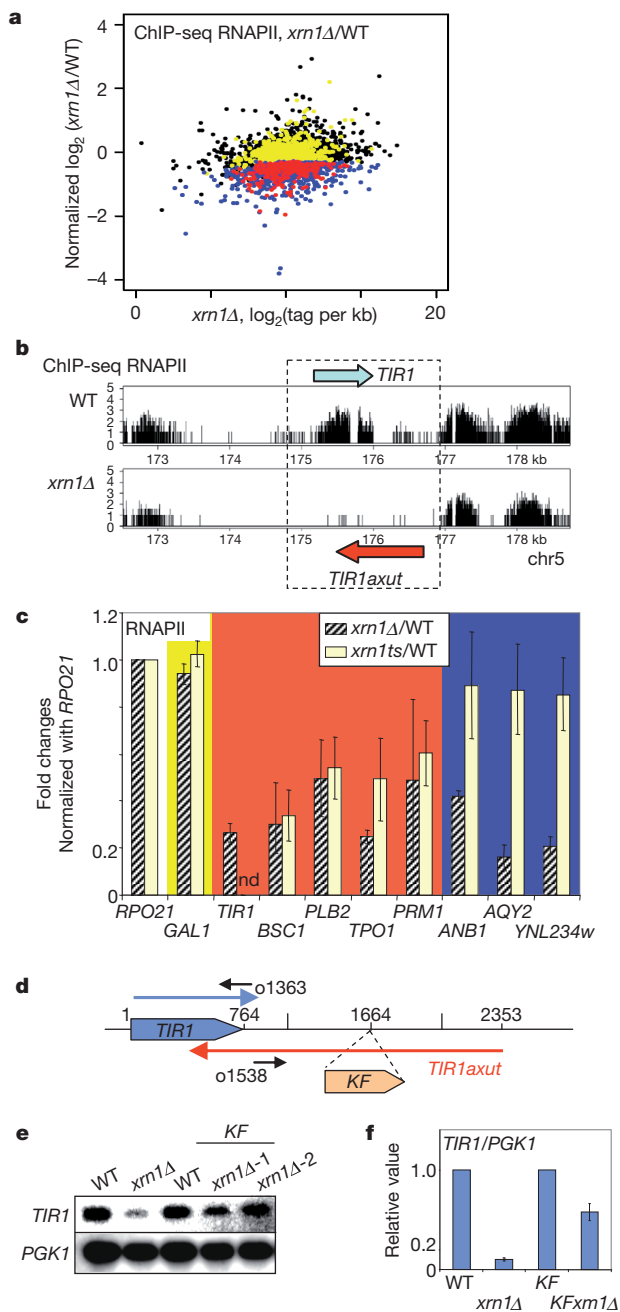


Figure 3 | Transcriptional gene silencing correlates with antisense XUT accumulation. **a**, ORFs with reduced tag densities in *xrn1Δ* strain compared to wild type by RNAPII ChIP-seq analysis ($P < 0.01$, Methods). Selected reduced ORFs associated with (class 1, red) or without antisense XUT (class 2, blue); ORFs with no significant variation with antisense XUT ($P > 0.01$, class 3, yellow) or without (class 4, black). **b**, ChIP-seq density profiles of *TIR1* locus in wild-type and *xrn1Δ* strains. **c**, RNAPII-ChIP analyses by quantitative PCR (mean \pm s.d., $n = 3$) using primers within class 4: *RPO21*; class 3: *GAL1*; class 1: *TIR1*, *TPO1*, *BSC1*, *PLB2*, *PRM1* and class 2: *ANB1*, *AYQ2*, *YNL234w*, using wild-type, *xrn1Δ* and *xrn1ts* strains. *xrn1Δ*/WT and *xrn1ts*(37 °C)/WT(37 °C) RNAPII ratios normalized by *RPO21*. *xrn1ts* ChIP was performed after 180 min at 37 °C. *TIR1* expression was undetectable upon heat shock (nd). Colours as in **a**. **d**, *TIR1* locus and insertion of the *KANMX* cassette. Probes to detect sense and antisense RNAs are indicated. **e**, *TIR1* RNA levels are directly controlled by *TIR1axut*. Northern-blot with polyA⁺ RNA using sense-specific probes to detect *TIR1* and *PGK1* mRNAs. Strains: WT, *xrn1Δ*, *KFWT* and *KF:xrn1Δ* (two clones). **f**, Quantification of three independent Northern blots using *PGK1* for normalization (mean \pm s.d., $n = 3$); respective wild-types were set to 1.

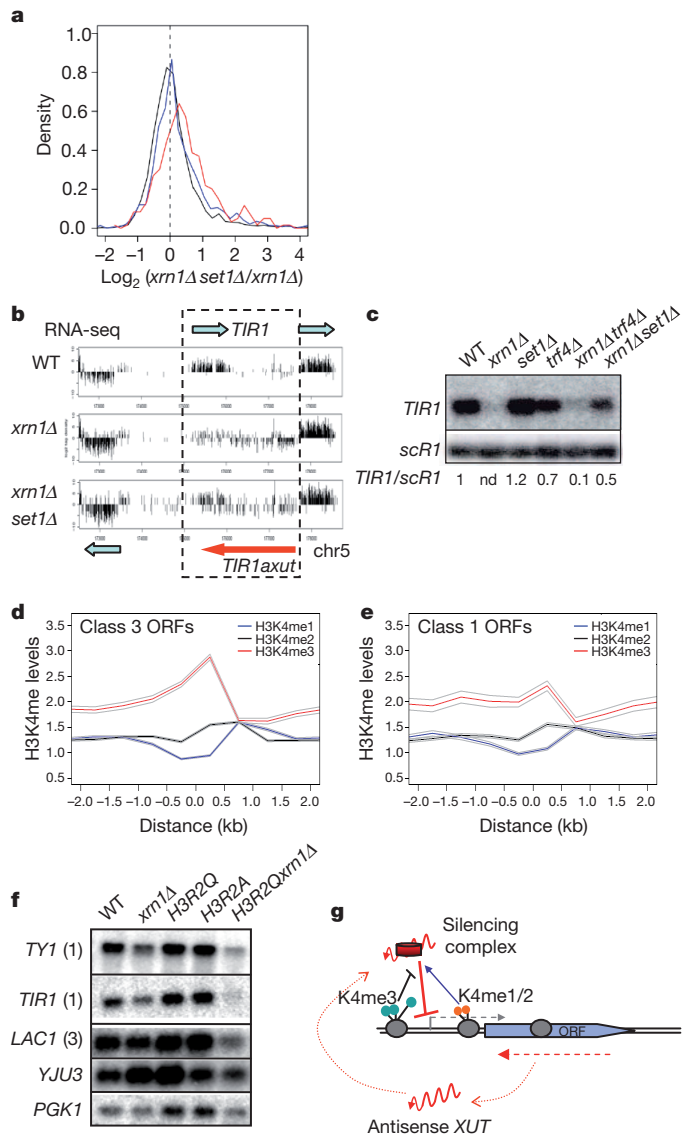


Figure 4 | Antisense XUTs mediate transcriptional gene silencing through Set1-dependent histone methylation. **a**, Repression of class 1 genes partly depends on Set1. Distributions of the \log_2 ratios of RNA-seq tag densities *xrn1set1Δ/xrn1Δ* for the ORFs; class 1, red; class 2, blue; others, black. **b**, *TIR1* locus transcriptomes in WT, *xrn1Δ* and *xrn1Δset1Δ* strains. Blue arrow, *TIR1* RNA; red arrow, *TIR1axut*; settings as in Fig. 1d. **c**, Northern blot validation of **b** with total RNAs. from indicated strains; probes as in Fig. 3f. **d**, **e**, Histone H3K4 methylation (mean \pm s.e.m.) for genes with antisense XUTs in class 3 ($n = 820$) and class 1 ($n = 273$), respectively. Data from wild type retrieved from previous publication²⁸ were plotted for H3K4me1 (blue), H3K4me2 (black) and H3K4me3 (red); abscissa, distance to start codon (kb). **f**, Northern blot of polyA⁺ RNA revealing genes with antisense XUTs: *TIR1*, *TY1* (class 1), *LAC1* (class 3), or without: *YJU3*, *PGK1* (class 4) for indicated strains. Note that *TIR1* and *TY1* mRNAs are less sensitive to *xrn1Δ* due to background differences. **g**, Model depicting XUT-mediated repressive transcriptional activity enhanced by H3K4me1/2 (orange) and antagonized by H3K4me3 (blue).

that class 1 genes are controlled by Xrn1, potentially through their respective antisense ncRNAs. Interruption of the *TIR1axut* in the *xrn1Δ* strain (Fig. 3d and Supplementary Fig. 15) re-established 70% of the *TIR1* mRNA level (Fig. 3e, f). Residual *TIR1axut* detection, due to either alternative initiation start sites of antisense ncRNAs or read-through over the inserted cassette, might explain the residual repressive activity on *TIR1*. Altogether, these data strongly support a direct role for antisense XUTs in gene silencing.

The ncRNA-mediated *TY1* silencing requires the Set1 activity¹² which represses basal transcription of many genes associated with antisense

transcripts²⁷. RNA-seq analyses performed in a *set1Δxrn1Δ* strain revealed that only the class 1 genes have higher RNA levels in the absence of Set1 (Fig. 4a; $P < 10^{-6}$, one-tailed two-sample Kolmogorov–Smirnov test), indicating that Set1 has a widespread role in antisense ncRNA-mediated gene repression. Remarkably, Set1 is not involved in XUT expression because the *set1Δxrn1Δ* and *xrn1Δ* strains showed similar levels of XUTs (Supplementary Fig. 16, $P = 0.2$). Furthermore, *TIR1* silencing is partly mediated by Set1-dependent histone H3 lysine 4 methylation (H3K4me) because disrupting the Set1 catalytic activity re-established high levels of *TIR1* RNA as shown in *xrn1Δset1Δ*, *xrn1Δset1G951S* and *xrn1ΔH3K4A* strains (Fig. 4b, c and Supplementary Fig. 17a, b). To dissect the Set1 role, we compared the H3K4me1, H3K4me2 and H3K4me3 levels²⁸ in the wild type over the genes with antisense XUTs. Class 3 genes show H3K4me1/2/3 patterns (Fig. 4d) equivalent to all ORFs (Supplementary Fig. 18). By contrast, class 1 genes present 60% reduction of H3K4me3 levels but normal H3K4me2/me1 patterns (Fig. 4e). This indicates that although they have similar RNAPII distribution as class 3 genes (Supplementary Fig. 19), class 1 genes are marked and/or sensitized to their respective antisense ncRNA by low H3K4me3 levels. A tantalizing hypothesis is that antisense ncRNA-mediated gene silencing requires H3K4me1 and/or me2 but is antagonized by H3K4me3. Our model predicts that class 1 but also class 3 genes would be further silenced in absence of H3K4me3. To test this possibility, we performed analyses in a strain specifically depleted for H3K4me3 (H3R2Q mutant), without disrupting the Set1 complex²⁹. In addition to a synergetic repression of *TY1* and *TIR1* mRNA (class 1 genes), we observed a decrease of the class 3 *LAC1* RNA levels in the *H3R2Qxrn1Δ* strain (Fig. 4f and Supplementary Fig. 20a). H3K4me3 is not involved in the expression of the respective antisense XUTs, similarly expressed in the *xrn1Δ* and *H3R2Qxrn1Δ* strains (Supplementary Fig. 20b). By contrast the *PGK1* and *YJU3* genes (class 4), both devoid of antisense XUTs, were not affected in *H3R2Qxrn1Δ*, supporting a model in which H3K4me3 antagonizes the negative activity of antisense ncRNA potentially associated with unknown silencing factors (Fig. 4g).

In conclusion, our results provide evidence for a novel class of cryptic antisense transcripts degraded by the cytoplasmic 5'–3' RNA decay pathway. We propose that Set1-dependent histone methylation has a key role in controlling the antisense ncRNA-repressive activity using H3K4me3 as a molecular switch. Given the broad class of XUTs, we already anticipate also a post-transcriptional role(s). Because the key components of this pathway are distributed among the eukaryotic kingdom, the mechanism is likely to be conserved throughout evolution.

METHODS SUMMARY

RNA-seq libraries were generated according to the manufacturer's instructions using ribo[−]/polyA⁺ RNAs extracted from *S. cerevisiae* strains. ChIP-seq was performed according to Illumina protocols and sequenced on GAIIx. Tag profiles were analysed using appropriated biocomputational and statistical approaches (Methods). Northern-blot and ChIP (using anti-Rpb1-CTD 8wg16 antibody) were performed using standard techniques (Methods). Sequence data are publicly available at NCBI Sequence Read Archive under accession number SRA030505 and at <http://vm-gb.curie.fr/XUT/index.htm>.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 31 December 2010; accepted 14 April 2011.

Published online 22 June 2011.

- Bernstein, E. & Allis, C. D. RNA meets chromatin. *Genes Dev.* **19**, 1635–1655 (2005).
- Moazed, D. Small RNAs in transcriptional gene silencing and genome defence. *Nature* **457**, 413–420 (2009).
- Swiezewski, S., Liu, F., Magusin, A. & Dean, C. Cold-induced silencing by long antisense transcripts of an *Arabidopsis* Polycomb target. *Nature* **462**, 799–802 (2009).
- Yu, W. *et al.* Epigenetic silencing of tumour suppressor gene *p15* by its antisense RNA. *Nature* **451**, 202–206 (2008).
- Yap, K. L. *et al.* Molecular interplay of the noncoding RNA *ANRIL* and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of *INK4a*. *Mol. Cell* **38**, 662–674 (2010).

- Huarte, M. & Rinn, J. L. Large non-coding RNAs: missing links in cancer? *Hum. Mol. Genet.* **19**, R152–R161 (2010).
- Chow, J. & Heard, E. X inactivation and the complexities of silencing a sex chromosome. *Curr. Opin. Cell Biol.* **21**, 359–366 (2009).
- Nagano, T. *et al.* The *Air* noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science* **322**, 1717–1720 (2008).
- Amaral, P. P., Dinger, M. E., Mercer, T. R. & Mattick, J. S. The eukaryotic genome as an RNA machine. *Science* **319**, 1787–1789 (2008).
- Berretta, J. & Morillon, A. Pervasive transcription constitutes a new level of eukaryotic genome regulation. *EMBO Rep.* **10**, 973–982 (2009).
- Camblong, J. *et al.* Trans-acting antisense RNAs mediate transcriptional gene cosuppression in *S. cerevisiae*. *Genes Dev.* **23**, 1534–1545 (2009).
- Berretta, J., Pinskaya, M. & Morillon, A. A cryptic unstable transcript mediates transcriptional trans-silencing of the Ty1 retrotransposon in *S. cerevisiae*. *Genes Dev.* **22**, 615–626 (2008).
- Matsuda, E. & Garfinkel, D. J. Posttranslational interference of Ty1 retrotransposition by antisense RNAs. *Proc. Natl Acad. Sci. USA* **106**, 15657–15662 (2009).
- Aravind, L., Watanabe, H., Lipman, D. J. & Koonin, E. V. Lineage-specific loss and divergence of functionally linked genes in eukaryotes. *Proc. Natl Acad. Sci. USA* **97**, 11319–11324 (2000).
- Jacquier, A. The complex eukaryotic transcriptome: unexpected pervasive transcription and novel small RNAs. *Nature Rev. Genet.* **10**, 833–844 (2009).
- Long, R. M. & McNally, M. T. mRNA decay: X (*XRN1*) marks the spot. *Mol. Cell* **11**, 1126–1128 (2003).
- Nagalakshmi, U. *et al.* The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* **320**, 1344–1349 (2008).
- Neil, H. *et al.* Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. *Nature* **457**, 1038–1042 (2009).
- Yassour, M. *et al.* Strand-specific RNA sequencing reveals extensive regulated long antisense transcripts that are conserved across yeast species. *Genome Biol.* **11**, R87 (2010).
- Chernyakov, I., Whipple, J. M., Kotelawala, L., Grayhack, E. J. & Phizicky, E. M. Degradation of several hypomodified mature tRNA species in *Saccharomyces cerevisiae* is mediated by Met22 and the 5'–3' exonucleases Rat1 and Xrn1. *Genes Dev.* **22**, 1369–1380 (2008).
- Fatica, A., Morlando, M. & Bozzoni, I. Yeast snoRNA accumulation relies on a cleavage-dependent/polyadenylation-independent 3'-processing apparatus. *EMBO J.* **19**, 6218–6229 (2000).
- Xu, Z. *et al.* Bidirectional promoters generate pervasive transcription in yeast. *Nature* **457**, 1033–1037 (2009).
- Thompson, D. M. & Parker, R. Cytoplasmic decay of intergenic transcripts in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **27**, 92–101 (2007).
- Wyers, F. *et al.* Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. *Cell* **121**, 725–737 (2005).
- Dichtl, B., Stevens, A. & Tollervey, D. Lithium toxicity in yeast is due to the inhibition of RNA processing enzymes. *EMBO J.* **16**, 7184–7195 (1997).
- Johnson, A. W. Rat1p and Xrn1p are functionally interchangeable exoribonucleases that are restricted to and required in the nucleus and cytoplasm, respectively. *Mol. Cell. Biol.* **17**, 6122–6130 (1997).
- Pinskaya, M. & Morillon, A. Histone H3 lysine 4 di-methylation: a novel mark for transcriptional fidelity? *Epigenetics* **4**, 302–306 (2009).
- Pokholok, D. K. *et al.* Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell* **122**, 517–527 (2005).
- Kirmizis, A. *et al.* Arginine methylation at histone H3R2 controls deposition of H3K4 trimethylation. *Nature* **449**, 928–932 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank B. Séraphin, L. Bénard and J. O'Sullivan for support and advice; B. Dichtl for insights into lithium treatment data normalization; M. Wéry and A. Taddei for helpful discussions; A. Johnson, T. Kouzarides and V. Géli for generous gift of plasmids and strains. Special thanks to M. Describes, C. Jubin and S. Lair for technical assistance. We thank L. Steinmetz and M. Chodder for sharing unpublished results. E.L.V.D. benefits from an FRM fellowship. This work has benefited from facilities and expertise of the IMAGiF sequencing platform (Centre de Recherche de Gif). This work was financially supported by the Canceropole Ile de France, the ANR “REGULncRNA” and ERC “EPIncRNA” starting grant.

Author Contributions E.L.V.D. performed molecular biology experiments, RNA-seq, ChIP-seq libraries and sequencing on the ILLUMINA platform. C.L.C., Y.D.-C., M.S. and C.T. performed statistical and bioinformatic analyses. S.G., M.K., V.R. and C.B. provided technical assistance to molecular biology experiments. A.M., C.T., E.L.V.D. and C.L.C. designed the experiments. P.L.-N. and S.L. performed RNA-seq libraries and NGS sequencing on the SOLiD platform; A.N. managed sequencing on the SOLiD platform. E.L.V.D., C.L.C. and A.N. contributed to the writing. C.T. and A.M. wrote the paper. C.T. and A.M. planned the project.

Author Information Sequence data are publicly available at NCBI Sequence Read Archive under accession number SRA030505 and at <http://vm-gb.curie.fr/XUT/> index.htm. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.M. (antonin.morillon@curie.fr) or C.T. (thermes@cgm.cnrs-gif.fr).

METHODS

Yeast strains and growth media. Strains used in this study are from the Euroscarf collection (S288C, BY4741 background), except for the RNAPII and histone H3 studies. For RNAPII, we used the strains *rpb1-1* and *rpb1-1xrn1Δ* (ref. 12). For histone H3 we used strains WT-H3, *H3K4A*, *xrn1Δ* and *H3K4Axrn1Δ* previously manipulated¹² and WT-H3, *H3R2A* and *H3R2Q* generously provided by T. Kouzarides²⁹ in which we deleted *XRN1* gene (this work). The list of BY strains is WT, *xrn1Δ*, *set1Δ*, *set1Δxrn1Δ*, *trf4Δ*, *xrn1Δtrf4Δ*, *set1G951S* and *set1G951Sxrn1Δ*. The *set1G951S* strain was kindly provided by V. Géli. Gene deletions were introduced by transformation of PCR fragments generated with specific primers (sequences can be obtained upon request) using the appropriate plasmids³⁰.

Growth media were prepared by standard methods using rich YPD media (yeast peptone dextrose adenine, Gibco) or minimal CSM media (MP biochemical) containing 2% glucose.

For lithium induction, cells were grown in CSM media and then transferred during the indicated time in CSM media with 100 mM lithium as indicated²⁵.

For *Xrn1* conditional mutant experiments, the *xrn1Δ* strain was transformed with a wild-type *XRN1* plasmid (PAM27, pAJ52) or *xrn1ts* plasmid (PAM143, pAJ53), both plasmids from a gift from A. Johnson. Transformed cells were grown in CSM-URA media with glucose at 30 °C. Heat shock was performed at 37 °C and cells collected at the indicated times.

RNA extraction, polyA⁺ and ribo⁻ RNA purifications. Total RNA was extracted using the hot phenol extraction procedure. PolyA⁺ RNAs were purified on oligodT Dynabeads (Invitrogen) and ribosomal RNA were depleted using the RiboMinus kit (Invitrogen). RNA quality was checked on agarose gel (Supplementary Fig. 2) and with the Bioanalyzer (Agilent) ensuring the quantification of RNA peaks recommended by the manufacturer.

Northern blotting. RNAs were loaded on denaturing 1% agarose gels containing formaldehyde and transferred to nitrocellulose membranes (Hybond XL). Membranes were cross-linked by ultraviolet irradiation and hybridized over-night at 65 °C with either ³²P-labelled DNA probes or oligonucleotide probes in PERFECT-HYB PLUS buffer (Sigma). Blots were washed at 65 °C for 10 min once with 2× SSC, 1% SDS and twice with 0.1× SSC, 1% SDS. DNA probes were obtained by random primed labelling (Stratagene) of specific DNA fragments generated by PCR. PCR primers are available upon request. *PGK1* and *scR1* RNAs were used as loading controls. Northern signals were quantified using ImageQuant software, normalized with *scR1* or *PGK1* intensities and wild-type ratios were set arbitrarily to 1. Error bars correspond to the standard deviation over three independent cultures.

Reverse transcription. Reverse transcription was performed according to the manufacturer's instructions (Invitrogen, SuperScript) on total RNA. Specific primers were used to amplify *TIR1axut*, *LAC1axut* and *scR1* loading control. Quantitative PCR were performed with the LightCycler 480 (Roche) using SYBR Green. PCR were normalized with *scR1* signals and wild-type ratios adjust to 1 for comparison. Error bars correspond to standard deviation over three independent reverse transcription reactions.

5' RACE analysis. 5' RACE experiments were done using a Firstchoice RLM-RACE kit from Ambion following the manufacturer's instructions. Total RNA from wild-type or *xrn1Δ* cells was used. Reverse transcription was performed with primers A, B, and C to detect the 5' ends of *SUC2uxut* (Supplementary Fig. 7). Subsequently, PCR was done using nested primers (sequences available upon request). PCR products were cloned and sequenced.

Chromatin immunoprecipitation. Chromatin immunoprecipitations were performed essentially as described previously¹². Yeast strains were grown to $D_{600} = 0.5$ in YPD at 30 °C, and cross-linked 20 min by the addition of formaldehyde to a final concentration of 1.2%. Crosslinked reaction was quenched by adding glycine at 0.5 M final concentration. Chromatin was sonicated to obtain 400–500 nucleotides DNA fragment and 200 µg of sonicated chromatin was immunoprecipitated for 3 h at 21 °C on Pan mouse Dynabeads (Invitrogen) coated with specific antibody against the carboxy-terminal domain of Rpb1 (8WG16, Millipore). All immunoprecipitations were repeated at least three times with different chromatin extracts. Immunoprecipitated DNA was quantified by real-time PCR using the LightCycler 480 (Roche) with primer pairs (sequences available upon request). Signals are expressed as percentage of input DNA relatively to *RPO21* (coordinate: chromosome 4: 210562 to 205361). Error bars correspond to standard deviations of three independent experiments.

Library preparation for RNA-seq. The extracted RNAs (polyA⁺ or ribo⁻) were submitted to RNA fragmentation as recommended in the ABI whole transcriptome library preparation kit. RNA adaptors were ligated before cDNA preparation to ensure strand recognition. Typically, DNAs ranging from 150 to 250 nucleotides were size-selected on polyacrylamide gel followed by PCR amplification (15 cycles). Denatured single-stranded DNA was then hybridized on beads before emulsion PCR and deposited on SOLiD slides for sequencing. Runs were performed on the

SOLiD V3 sequencing machine and 50 nucleotides single reads were generated on colour code format.

For Illumina sequencing, RNAs were subjected to sequencing on the Genome Analyzer IIX sequencing machine and 38 nucleotides reads were generated. The same polyA⁺ samples as for the SOLiD libraries (but with Illumina adapters added by PCR) were used to confirm that sequencing technologies do not introduce bias in the reads. The duplicates libraries for wild type and *xrn1Δ*, lithium and *ts* were sequenced with Illumina technology (Supplementary Table 1).

Library preparation for ChIP-seq. For ChIP-seq, chromatin immunoprecipitation was done as described above, the only modification being that reactions were scaled up to obtain sufficient amounts of material for library preparation. Instead of 200 µg, 800 µg of sonicated chromatin was used. The sequences were obtained on Genome Analyzer IIX sequencer with 38 nucleotides reads (Supplementary Table 1).

Sequencing data. For SOLiD sequencing, 50-nucleotide sequence reads were identified using the standard SOLiD base-calling software and then aligned to the reference genome (*Saccharomyces cerevisiae* S288c retrieved from SGD, <http://www.yeastgenome.org/>) using Mapreads (v2.4.1) software allowing up to six mismatches. For Illumina/Solexa sequencing, 38-nucleotide sequence reads were identified using the standard Illumina base-calling software and then aligned to the reference genome using the ELAND (CASAVA pipeline) software allowing up to two mismatches within the 32-nucleotide seed. For multiplex samples, reads were partitioned and those lacking an intact barcode or index were discarded. Replicate experiments for ribo⁻ samples produced read density values that strongly correlated to each other (Supplementary Fig. 4a, Pearson $R = 0.96$, $P < 10^{-15}$). Sequencing data obtained with the polyA⁺/WT RNA sample using either the SOLiD or the Illumina/Solexa device were highly correlated to each other (Supplementary Fig. 4b, $R = 0.98$, $P < 10^{-15}$).

Transcriptome analysis and normalization. Genome was annotated according to SGD (<http://www.yeastgenome.org/>, 5 January 2010); SUT (stable unannotated transcript) and CUT (cryptic unstable transcript) annotations were retrieved²². Tag densities were computed for each transcript in each sample and in all cases, except for tRNAs, only tags mapping to unique positions were considered. Tag densities for tRNAs were computed (for each anticodon) by using the tags mapping to multiple positions. Except for the lithium and *set1Δ* transcriptomes, the tag densities were normalized in such a way that snoRNAs and tRNAs present the same levels in the wild-type and *xrn1Δ* strains (Supplementary Figure 3). The candidate transcripts obtained by the segmentation process that presented an *xrn1Δ*/WT ratio significantly larger than the background ratio were retained as XUTs. To determine whether the expression level of a gene (ORF) was significantly different in the *xrn1Δ* and wild-type strains, a P -value was computed based on negative binomial distribution (edgeR package³¹ from Bioconductor, <http://www.bioconductor.org/>); for this comparison, individual ORF expression levels in the *xrn1Δ* and wild-type strains were compared to the ratio of the corresponding mean ORF expression values; transcripts with a P -value < 0.01 were retained. In the wild-type strain, a fraction of XUTs presents expression levels close to the background level of the density profile; to compare XUT expression levels in the ribo⁻ and polyA⁺ samples, only those presenting levels 50% higher than the background level of the corresponding profile were retained.

An extensive study was performed to identify the best way to normalize the RNA-seq values in the presence of lithium. As suggested by previous studies²⁵, tRNAs and snoRNAs levels are affected by lithium addition with mean ratios of 1.8 and 2.1, respectively (Supplementary Fig. 11a–d). In contrast, ribosomal proteins genes-derived transcripts present only a slight increase (1.5 fold) in the *xrn1Δ* strain, and are insensitive to lithium addition (mean ratio of 1.1, Supplementary Fig. 11e, f). The mean ribosomal proteins genes transcript value was thus used as reference to normalize RNA-seq values in the presence of lithium.

For the *set1Δ* transcriptome, levels were normalized by the mean level of all ORFs taken as reference, because the Set1 complex controls snoRNA transcription³².

Segmentation of massive sequencing data. The RNA-seq procedure generates considerable variation of tag coverage along genes complicating the segmentation process. To address this problem, we designed a crude heuristic to detect candidate transcripts. Segmentation was performed on massive sequencing data obtained from the ribo⁻ samples extracted from the *xrn1Δ* strain. The tag profile was built by computing the tag density within a sliding 120 base pairs (bp) window, using a 1-bp step size; segments were delimited on the tag profile using a threshold of 0.06 tag per nucleotide (50% of the annotated CUTs present a tag density lower than this threshold). To define transcripts, the extremities of each segment were adjusted to the first and last internal tags. Segments overlapping genes annotated on the same strand (ORFs, stable RNAs) were then removed. The remaining consecutive segments of similar densities (ratio smaller than twofold) distant by less than 250 bp were joined (only the 5' end of the RNA pieces obtained by fragmentation of the initial transcripts was sequenced); the resulting segments longer than 250 bp were retained as final transcripts corresponding to 543

SUTs, 183 CUTs and 932 new transcripts (New). The extremities of XUTs tested experimentally were manually repositioned according to experimental data. With this procedure, previously defined SUTs were covered at 87% by the detected segments and the segments intersecting the SUTs were covered at 80% by the SUTs; in addition, 90% of the SUTs were covered by a unique detected segment (Supplementary Fig. 5).

XUTs are polyadenylated and RNAPII dependent. Among the 1,658 XUTs, 683 were detected at low levels in the wild-type ribosome sample. Interestingly, 493 (72%) were also detected within the polyA⁺ fraction (Supplementary Fig. 8a, b), strongly supporting that XUTs are polyadenylated, as shown for *PHO84*, *GAL10*- and *TY1*-associated ncRNAs^{12,33,34}. To confirm that XUTs are synthesized by RNAPII, we monitored the steady-state levels of the *SUC2uxut1* and *SUC2uxut3* mRNAs upon depletion of Rpb1, the main subunit of RNAPII (ref. 35). The *SUC2uxut1* mRNA levels were significantly reduced after 60 min of RNAPII inactivation similarly to the class 2 *ACT1* mRNA (Supplementary Fig. 8c, d). In contrast, the levels of *scR1* and rRNA, transcribed by RNA polymerase III and I, respectively, remained constant. Altogether these data support that XUTs RNA are polyadenylated and synthesized by the RNA polymerase II.

Analysis of *xrn1* conditional mutant. To examine possible late secondary effects of *XRN1* deletion, we asked whether a conditional mutant of Xrn1 would rapidly accumulate XUTs after Xrn1 inactivation and show repression of the same group of genes defined in Fig. 3a (class 1). To answer this question, we determined the kinetics of the ncRNA-mediated Ty1 regulation using an *xrn1* thermosensitive (ts) mutant²⁶. Interestingly, *TY1* ncRNA accumulated around 30 min after heat shock (Supplementary Figure 14a). In addition, *TY1* ncRNA accumulation preceded the decrease of the *TY1* mRNA starting approximately at 60 min, confirming the causal link between XUT accumulation and gene repression^{12,13}. Interestingly, the *TY1* ncRNA kinetics reflect other XUTs dynamic because *SUC2uxut1*, *SUC2uxut3* and *TPO1axut* also appeared at 30 min, indicating a rapid and general

accumulation of the XUTs upon Xrn1 inactivation. Indeed, RNA-seq experiments showed that 75% of the XUTs detected in *xrn1Δ* strain accumulated similarly after 90 min of Xrn1 inactivation (Supplementary Fig. 14b–d), indicating that XUTs are primary targets of Xrn1 RNA decay pathway. To determine whether the regulation of class 1 genes is a late or early event, we performed RNAPII-ChIP experiments in the *xrn1Δ* and *xrn1ts* strains (after 180 min in 37 °C) on a selection of genes belonging to the different classes defined in Figure 3a. Strikingly, the class 1 genes showed a similar decrease of RNAPII occupancy in the two *XRN1*-defective strains (Figure 3c), indicating that Xrn1 inactivation has an early effect on RNAPII on those genes. By contrast, the class 2 genes are only suppressed in *xrn1Δ* strain supporting a late and likely indirect effect of Xrn1 inactivation on those genes. We conclude that the class 1 genes are directly controlled by Xrn1, potentially through their respective antisense ncRNAs.

30. Longtine, M. S. *et al.* Additional modules for versatile and economical PCR-based gene deletion and modification in *Saccharomyces cerevisiae*. *Yeast* **14**, 953–961 (1998).
31. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression. *Bioinformatics* **26**, 139–140 (2010).
32. Dichtl, B., Aasland, R. & Keller, W. Functions for *S. cerevisiae* Swd2p in 3' end formation of specific mRNAs and snoRNAs and global histone 3 lysine 4 methylation. *RNA* **10**, 965–977 (2004).
33. Pinskaya, M., Gourvennec, S. & Morillon, A. H3 lysine 4 di- and tri-methylation deposited by cryptic transcription attenuates promoter activation. *EMBO J.* **28**, 1697–1707 (2009).
34. Camblong, J., Iglesias, N., Fickentscher, C., Dieppois, G. & Stutz, F. Antisense RNA stabilization induces transcriptional gene silencing via histone deacetylation in *S. cerevisiae*. *Cell* **131**, 706–717 (2007).
35. Nonet, M., Scafe, C., Sexton, J. & Young, R. Eucaryotic RNA polymerase conditional mutant that rapidly ceases mRNA synthesis. *Mol. Cell. Biol.* **7**, 1602–1611 (1987).

The ribosome uses two active mechanisms to unwind messenger RNA during translation

Xiaohui Qu^{1,2}, Jin-Der Wen^{1,2,†}, Laura Lancaster³, Harry F. Noller³, Carlos Bustamante^{1,2,4} & Ignacio Tinoco Jr²

The ribosome translates the genetic information encoded in messenger RNA into protein. Folded structures in the coding region of an mRNA represent a kinetic barrier that lowers the peptide elongation rate, as the ribosome must disrupt structures it encounters in the mRNA at its entry site to allow translocation to the next codon. Such structures are exploited by the cell to create diverse strategies for translation regulation, such as programmed frameshifting^{1,2}, the modulation of protein expression levels^{3,4}, ribosome localization⁵ and co-translational protein folding⁶. Although strand separation activity is inherent to the ribosome, requiring no exogenous helicases⁷, its mechanism is still unknown. Here, using a single-molecule optical tweezers assay on mRNA hairpins, we find that the translation rate of identical codons at the decoding centre is greatly influenced by the GC content of folded structures at the mRNA entry site. Furthermore, force applied to the ends of the hairpin to favour its unfolding significantly speeds translation. Quantitative analysis of the force dependence of its helicase activity reveals that the ribosome, unlike previously studied helicases, uses two distinct active mechanisms to unwind mRNA structure: it destabilizes the helical junction at the mRNA entry site by biasing its thermal fluctuations towards the open state, increasing the probability of the ribosome translocating unhindered; and it mechanically pulls apart the mRNA single strands of the closed junction during the conformational changes that accompany ribosome translocation. The second of these mechanisms ensures a minimal basal rate of translation in the cell; specialized, mechanically stable structures are required to stall the ribosome temporarily^{1,2}. Our results establish a quantitative mechanical basis for understanding the mechanism of regulation of the elongation rate of translation by structured mRNAs.

To monitor ribosome translation and mRNA unwinding with optical tweezers, we use a hairpin-forming mRNA which has a single-stranded 5' overhang that includes the Shine–Dalgarno sequence and the start codon for ribosome loading⁸ (Fig. 1a and Supplementary Fig. 1). Translation starts with the introduction of a translation mixture in the tweezers flow chamber.

Single-stranded mRNA enters the ribosome through the ring-shaped mRNA entry site formed by ribosomal proteins S3, S4 and S5, and wraps around the neck of the 30S ribosomal subunit in an RNA-rich channel that accommodates about 30 nucleotides⁹. The average tunnel diameter is less than that of an RNA double helix, so secondary structures in the mRNA must be disrupted for it to move through the tunnel. A crystal structure⁹, bulk oligonucleotide displacement assays⁷ and the optical tweezers measurements reported here (Supplementary Fig. 2) yielded a distance of 13 ± 2 (s.d.) nucleotides from the first nucleotide in the peptidyl site to the mRNA entry site. Therefore, as shown in Fig. 2a, when the ribosome translates codon *i* at the aminoacyl site, translocation to the next codon requires the unwinding of codon *i* + 4 downstream.

To investigate the effect of hairpin stability on the rate of translation, we designed two hairpin-containing mRNAs, hpVal_{GC50} and

hpVal_{GC100}, that differ only in the hairpin region, specifically in the GC content of the codons downstream of ten Val codons (Fig. 2a and Supplementary Fig. 1). Translation of the third to the tenth Val codons on hpVal_{GC50} mRNA is accompanied by unwinding of a helix with 46% GC content. For hpVal_{GC100} mRNA, translation of the third to the sixth Val codons results in unwinding of the seventh to the tenth Val codons, which is a helix with similar GC content (42%); but translation of the seventh to the tenth Val codons is accompanied by unwinding of the first four codons (Arg or Ala) that follow the sequence of Val codons, a helix with 100% GC content. Translation steps preceding the third Val codon were excluded from data analysis because the finite mixing time in the flow chamber made rate measurements for these first codons difficult. Unless noted otherwise, elongation factors were used at saturation concentrations (Supplementary Fig. 3) so that their binding to the ribosome was not rate limiting. Under our conditions, less than 5% of the ribosomes stalled before translating all the Val codons in the hpVal_{GC50} mRNA.

To characterize helicase activity, we applied different constant forces to the ends of these hairpins. Figure 1b shows a typical translation trajectory revealing periods of constant mRNA end-to-end distance (pauses) followed by periods of sudden distance increase (bursts). The burst between two pauses represents ribosome translocation, that is, the movement of the ribosome from one codon to the next. We obtain single-codon resolution at forces greater than 11 pN; the upper force value accessible to these experiments is approximately 3 pN below the opening force, F_c , for the hairpin. When held at constant forces closer

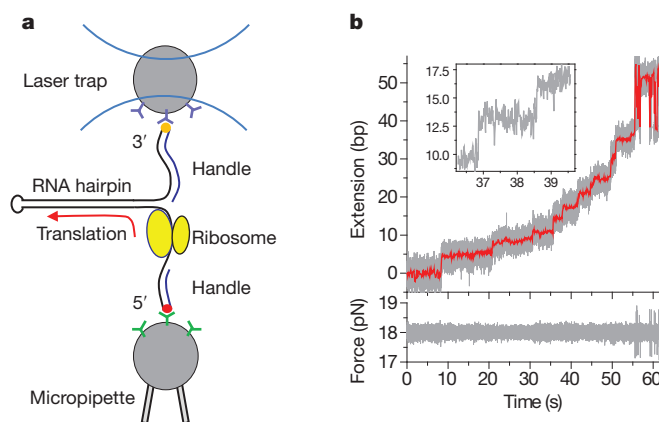


Figure 1 | Experimental set-up. **a**, Diagram showing the attachment of the ends of the mRNA hairpin to the two beads⁸. The mRNA sequences are shown in Supplementary Fig. 1. **b**, Trajectory for translation of hpVal_{GC50} mRNA under 18-pN force (grey, 1,000 Hz; red, 10 Hz). Extension change is in units of number of base pairs (bp) opened. Hopping of the residual hairpin is observed at the end of the trajectory owing to its instability under force. Inset, magnified view of stepwise extension change for two steps (100 Hz).

¹Jason L. Choy Laboratory of Single Molecule Biophysics and QRB Institute, University of California, Berkeley, California 94720, USA. ²Department of Chemistry, University of California, Berkeley, California 94720, USA. ³Department of Molecular, Cell and Developmental Biology and Center for Molecular Biology of RNA, University of California, Santa Cruz, California 95064, USA. ⁴Department of Molecular and Cellular Biology, Department of Physics and Howard Hughes Medical Institute, University of California, Berkeley, California 94720, USA. [†]Present address: Institute of Molecular and Cellular Biology, National Taiwan University, Taipei 10617, Taiwan.

to F_c , the hairpin unfolds spontaneously before translation is finished. To extend this range, and to measure the average translation rate for each individual ribosome down to 3 pN, we use a novel 'force drop/force jump' technique (Supplementary Fig. 4).

Burst times (Supplementary Fig. 5) have an average duration of 22 ± 3 ms (s.e.m.), and show no dependence on force or on the concentrations of elongation factors. Translation rates, which are determined by the pause lengths, depend monotonically on force (Fig. 2b)—under saturating elongation factor concentrations, the average pause time varies between ~ 2 s (at high force) and ~ 4 s (at low force)—and have a Michaelis–Menten dependence on the concentrations of elongation factors EF-G and EF-Tu (Supplementary Fig. 3). These observations suggest that translocation can be considered a single-barrier crossing process (Supplementary Discussion, section 1). A mechanochemical analysis¹⁰ of the dependence of the translation rate on elongation factor concentration at high and low force was done to determine which kinetic steps in the reaction are affected by the application of force on the junction and, thus, which kinetic steps involve motion of the ribosome through the junction. This analysis (Supplementary Discussion, section 2) showed that although the Michaelis constant, K_m , and maximum velocity, V_{max} , for both elongation factors increase as the force applied to the hairpin increases, their

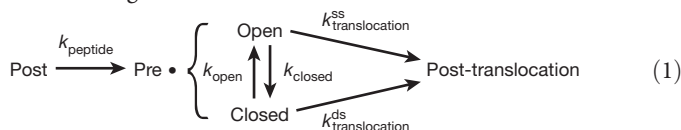
ratio is invariant. This invariance indicates that force does not affect the binding steps of these factors. Therefore, the translocation step, which does depend on force, can be separated from the previous biochemical steps (see scheme (1), below).

The translation rates for both hpVal_{GC50} mRNA ($\sim 50\%$ GC unwinding) and hpVal_{GC100} mRNA (100% GC unwinding) show a sigmoid dependence on force (Fig. 2b). It follows that the secondary structure of the mRNA represents a rate-limiting barrier to translation by the ribosome. The low-force plateau provides an estimate of the translation rate for double-stranded mRNA; the high-force plateau provides an estimate of the translation rate for single-stranded mRNA. The identical unwinding rates at the high-force plateaus for both 100% GC content and 50% GC content indicate that single-strand rates are not significantly affected by the downstream sequences. The force, F_M , at the midpoint of the sigmoid transition between double-strand and single-strand rates increases with increasing GC content, consistent with the greater thermodynamic stability of GC-rich double strands. As expected, the average translation rates of unwinding the $\sim 50\%$ GC region of hpVal_{GC100} mRNA and of unwinding the hpVal_{GC50} mRNA agree at all forces, within experimental error (Supplementary Fig. 6).

Various qualitative unwinding models, such as a strand-exclusion mechanism and a helix-destabilizing mechanism¹¹, have been proposed for processive helicases on the basis of their structures. Mechanical studies of the unwinding activity of several nucleic-acid helicases^{12–14} and HIV-1 reverse transcriptase on DNA templates¹⁵ have been interpreted in terms of the quantitative Betterton model^{12,16}. The junction at the end of a double strand fluctuates rapidly between open and closed states owing to thermal energy. The Betterton model postulates that a helicase translocates through the open state of the junction and that an active helicase lowers the free energy of the open state relative to the closed state of the junction by the amount ΔG_d , thus biasing junction thermal fluctuations towards the open conformation. In a totally active helicase, ΔG_d is much greater than the base-pair free energy, so the junction is always open and no longer hinders translocation. A passive helicase ($\Delta G_d = 0$) depends on junction opening by thermal fluctuations to translocate. In general a helicase will show an unwinding activity between the two extremes.

Applying the Betterton model to determine the extent of destabilization, ΔG_d , of a helicase can be ambiguous when the step size is unknown and when backtracking can occur¹⁴. However, in our experiments, a three-base-pair step size is directly measured, and ribosome reverse translocation is not observed. Thus, our data and these two constraints are sufficient to rule out the Betterton model, as no combination of parameters in that model can simultaneously fit the observed low-force plateau, high-force plateau and midpoint force, F_M , of the ribosomal helicase activity (Fig. 2b). Additional interactions need to be incorporated to model the mRNA unwinding by a translating ribosome.

In the Betterton model, the value of ΔG_d that fits the midpoint force predicts a low-force plateau of nearly zero rate (Fig. 2b, dashed black line), which is much smaller than we observe. This indicates that in addition to biasing the junction thermal breathing, the ribosome has an active mechanism that directly breaks open a closed junction during translocation. A minimal kinetic mechanism required by our results is the following:



Here all biochemical steps in a translation cycle^{17,18} other than the actual translocation step are combined into a single irreversible step with rate constant k_{peptide} , whereas $k_{\text{translocation}}^{\text{ss}}$ and $k_{\text{translocation}}^{\text{ds}}$ are the rates of translocation through an open (single-stranded) and closed (double-stranded) junction, respectively. In scheme (1), k_{open} and

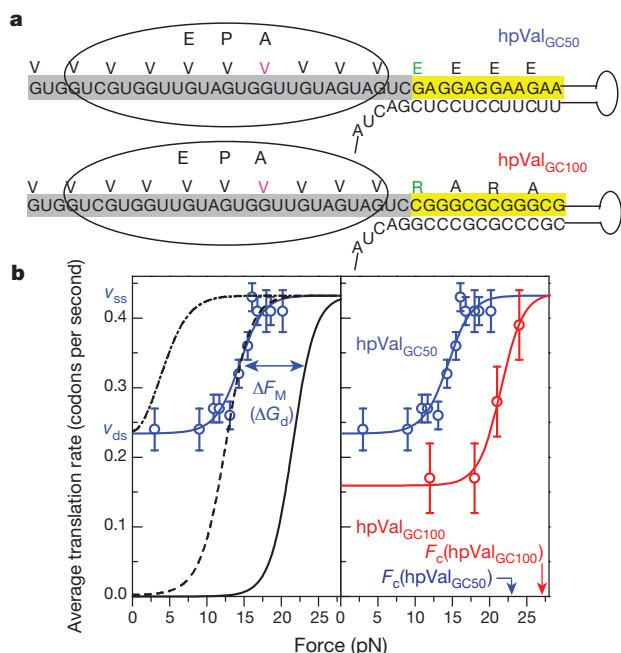


Figure 2 | Dependence of translation rate on force and mRNA GC content.

a, When the i th codon in the aminoacyl site (magenta) is translated, the subsequent translocation corresponds to unwinding the $(i + 4)$ th codon downstream (green) because the ribosome covers about 13 bases of single-stranded mRNA from its peptidyl site to the mRNA entry site. Only the sequence in the hairpin region is shown. E, exit site; P, peptidyl site; A, aminoacyl site. **b**, Left: translation-rate dependence on force for hpVal_{GC50} mRNA ($\sim 50\%$ GC unwinding). Blue circles show experimental data. Black solid, dashed and dot-dash lines show the force dependence predicted by the Betterton model, $v = v_{ss}/f_{open}$, with $\Delta G_d = 0, 1.1$ and 2.2 kcal mol⁻¹ bp⁻¹, respectively. The three lines represent a totally passive helicase (solid), the best fit to F_M and the high-force plateau (dashed), and the best fit to the two plateaux (dot-dash). The blue line shows the best fit to equation (1). The three fitting parameters, v_{ss} , v_{ds} and ΔG_d , largely determine the high-force plateau, the low-force plateau and ΔF_M (the shift in F_M relative to a totally passive helicase), respectively. Right: translation rate dependence on force for hpVal_{GC100} mRNA with 100% GC unwinding (red circles) and the best fit to equation (1) (red line). The data and best fit for hpVal_{GC50} mRNA (blue; from left-hand plots) is shown again for reference. The fitting results for both mRNAs are summarized in Supplementary Table 1. $n = 39$ –120 ribosomes for hpVal_{GC50} and 16–29 ribosomes for hpVal_{GC100}, at each force. Error bars, s.e.m.

k_{closed} depend on force, junction GC content and ΔG_d ; $k_{\text{translocation}}^{\text{ds}}$ depends only on junction GC content; and $k_{\text{translocation}}^{\text{ss}}$ is independent of these three factors. Setting $k_{\text{translocation}}^{\text{ds}} = 0$ reduces this kinetic mechanism to the Betterton model.

The kinetics equation derived from the mechanism of scheme (1) is given by (Supplementary Discussion, section 3):

$$v(F) = v_{\text{ss}} f_{\text{open}}(F) + v_{\text{ds}} (1 - f_{\text{open}}(F)) \quad (1)$$

where $v_{\text{ss}} = k_{\text{translocation}}^{\text{ss}} / (1 + k_{\text{translocation}}^{\text{ss}} / k_{\text{peptide}})$ and $v_{\text{ds}} = k_{\text{translocation}}^{\text{ds}} / (1 + k_{\text{translocation}}^{\text{ds}} / k_{\text{peptide}})$. Here $v(F)$ is the overall translation rate under force F , $f_{\text{open}}(F)$ is the probability that the junction is open at force F , $v_{\text{ds}} = v(f_{\text{open}} = 0)$ is the rate of ribosome translation through a closed junction and $v_{\text{ss}} = v(f_{\text{open}} = 1)$ is the rate of ribosome translation through an open junction.

The probability that the junction is open, $f_{\text{open}}(F)$, depends on the number of base pairs opened per translocation step (here three) and is quantified by the thermodynamic stability of the base pairs, ΔG_{bp} . This probability also depends on the effect of force applied to the ends of the hairpin, quantified by the energy ΔG_F , and on the destabilization of the junction induced by the ribosome, quantified by ΔG_d . Under our experimental conditions, the translocation step is rate-limiting for translation (Supplementary Discussion, section 4, and Supplementary Fig. 7); therefore

$$f_{\text{open}}(F) = \frac{1}{1 + \exp[(\Delta G_{\text{bp}} + \Delta G_F - \Delta G_d) / k_B T]} \quad (2)$$

where k_B is Boltzmann's constant and T is the temperature. We used MFOLD¹⁹ to calculate ΔG_{bp} ; the values give good agreement with our experimental results on folding/unfolding RNA with optical tweezers (Supplementary Fig. 8). We calculated ΔG_F using the worm-like chain expression for force versus extension²⁰. The parameters fitted to the experimental results are ΔG_d , v_{ss} and v_{ds} . The 50% and 100% GC sequence unwindings were fitted independently and yielded the same values of ΔG_d , 0.9 kcal mol⁻¹ (1.5 $k_B T$) per base pair, and v_{ss} , 0.43 or 0.44 codons per second (Fig. 2b and Supplementary Table 2). This value of ΔG_d is in the same range as those determined for the T7 DNA helicase¹² and HIV-1 reverse transcriptase¹⁵. The best-fit values of v_{ds} are 0.23 and 0.16 codons per second for the unwinding of 50% and 100% GC-containing hairpins, respectively.

Therefore, our data indicate that the ribosome uses two active mechanisms to promote junction unwinding: open-state stabilization (the role traditionally described for active helicases in the Betterton model, characterized by ΔG_d) and mechanical unwinding (a new active mechanism in which the ribosome translocates by applying force to the closed state of the junction, characterized by v_{ds}). In the first mechanism, the ribosome binds and interacts preferentially with the open form of the junction, favouring this state in the thermal fluctuations (corresponding to increasing $k_{\text{open}}/k_{\text{closed}}$ in scheme (1)) and therefore increasing the probability of the ribosome translocating unhindered. It has been shown that for nucleic-acid helicases this mechanism does not require an external energy source; the increase in unwinding rate is a result of the affinity between the helicase and the junction²¹. No known nucleic-acid helicase motifs are found in ribosomal proteins; however, a mutational study implicated several positively charged residues on ribosomal proteins S3 and S4 at the mRNA entry site in ribosome helicase activity. We assume that these groups preferentially interact with phosphate groups on the single-stranded mRNA backbone⁷. The identical values of ΔG_d extracted from unwinding of the 50% and 100% GC-containing hairpins indicate that the open-state stabilization mechanism in ribosomes has no significant base preference.

In the Betterton model, the ribosome can only translocate when encountering an open-state junction, whether occurring naturally or induced by the ribosome. However, when the ribosome encounters the junction in the closed state, the mechanical unwinding mechanism

comes into play, allowing the ribosome to break open the junction and translocate forward. A translating ribosome can thus use part of the energy of GTP hydrolysis and peptide bond formation, coupled to its translocation, to apply force and open the junction. The force applied by the ribosome on the junction lowers the free-energy barrier between the pre-translocation closed state and the post-translocation state, thereby opening this reaction path (labelled by $k_{\text{translocation}}^{\text{ds}}$ in scheme (1)). The fact that the value of v_{ds} for unwinding 100% GC is less than the value for unwinding 50% GC indicates that increasing GC content increases this barrier.

The molecular mechanism of ribosome translocation along mRNA is of great current interest²². Cryo-electron microscopy studies of ribosome translocation intermediates stalled by a pseudoknot have revealed a distorted conformation of the peptidyl-site transfer RNA, suggesting tension on the tRNA•mRNA linkage¹. Our results provide direct evidence that such tension could be used to open the RNA junction. Thus, we postulate that inter- and intrasubunit ribosomal conformational changes associated with translocation^{22–27} generate a force that pulls on the tRNA•mRNA complex and promotes unwinding at the mRNA entry site (Fig. 3). This tension may mediate ribosomal sensing of a high translocation barrier from a stable downstream mRNA structure and give rise to the observed strong correlation between frameshifting efficiency and pseudoknot mechanical stability²⁸. Single-molecule studies²⁹ have found that the rupture force between a ribosome and an mRNA goes from essentially zero with no tRNAs bound to 15 pN with tRNAs in both the peptidyl and the aminoacyl sites; this high mechanical stability of the tRNA•mRNA linkage is needed to communicate mechanical tension to the base-paired hairpin at the mRNA entry site during translocation.

The unwinding mechanisms revealed in our studies serve as a quantitative basis for understanding how elongation rates are regulated by structured mRNAs. The physiological consequences of a second, active mechanism for the ribosomal helicase are as follows. First, the existence of a mechanism that uses translocation energy to disrupt the

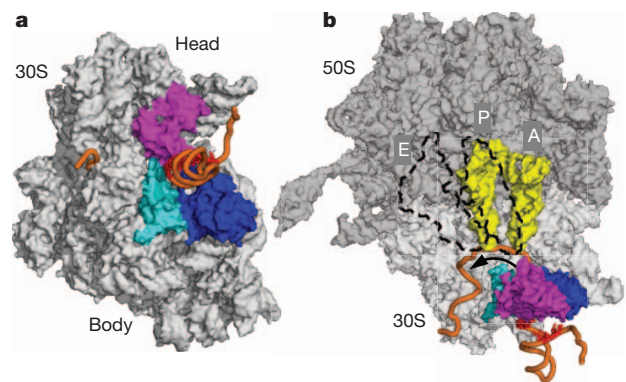


Figure 3 | The molecular arrangement of a translocating ribosome.

a, Structure of the 30S subunit viewed from the mRNA entry site. About 30 nucleotides of single-stranded mRNA (orange) enter the ring-shaped mRNA entry site formed by the three coloured ribosomal proteins, and wrap around the neck domain between the head and body. Secondary structures of the mRNA are excluded from the ribosome. **b**, Proposed ribosome unwinding mechanisms. The ribosome is rendered transparent so that the mRNA and the tRNAs inside can be clearly seen. Several base pairs (red) are destabilized by the ribosomal proteins at the entry site, which biases junction thermal fluctuations towards opening. Before translocation, the tRNAs (yellow) shift to the hybrid state (dashed). Then, driven by ribosomal conformational changes, the tRNA anticodons translocate from the peptidyl and aminoacyl sites on the 30S subunit to the exit and peptidyl sites, and pull the mRNA along (black arrow). When encountering an open junction, translocation rectifies the junction opening events. On encountering a closed junction, the pulling force on the mRNA breaks open the mRNA junction at the entry site. The 50S and 30S subunits are the Protein Data Bank entries with the IDs 2AW4 and 2AVY, respectively, and are modelled using PYMOL. The tRNAs and mRNA are for illustration purposes only.

double-stranded junction ensures a minimal basal rate of translation in the cell. It should also help to remove potential obstacles to elongation such as mRNA-binding proteins, tRNAs and so on. Second, the existence of such a mechanism means that no regular secondary structure represents an insurmountable barrier for the ribosome. Instead, specialized, mechanically stable structures, such as pseudoknots¹, or combinations of such structures (hairpins and internal Shine–Dalgarno sequences²) are used to stall the ribosome temporarily, thus providing a mechanical basis for cellular regulation of elongation rates. The existence of this additional mechanical unwinding mechanism in ribosomes suggests that it may also be used by other helicases, especially those with interdomain or intersubunit motions. The proposed unwinding model gives testable predictions for ribosomes bearing mutations in ribosomal proteins at the mRNA entry site or in proteins critical for translocation³⁰. Such studies will help elucidate the complex underlying molecular interactions involved in ribosome function and in the two active unwinding mechanisms established here.

METHODS SUMMARY

The materials, protocols and experimental procedures have been described in our previous study⁸, except that an optical tweezers instrument with higher spatial resolution was used for this study. All data were obtained at 24.5–26.5 °C.

Translation steps in each trajectory were detected by algorithms written in MATLAB. The pause times from all ribosomes under the same experimental condition were pooled together to calculate the average translation rate by a bootstrapping method: (i) half of all pause times are randomly selected, made into histograms and fitted to a single exponential; (ii) this procedure is repeated >1,000 times; and (iii) the average translation rate and its error are determined as the average and the s.d./ $\sqrt{2}$, respectively, of the fitted rate constants from all the repeats. This procedure is very insensitive to the bin size used.

In equation (2), the effect of force on the free energy of the open junction, ΔG_F , is calculated from the worm-like chain model²⁰ for RNA single strands as $\Delta G_F = -\int_0^F x_{\text{int}}(F') dF'$, where $x_{\text{int}}(F')$ is the total extension of the single strands released per translocation step under force F' .

Received 1 December 2010; accepted 15 April 2011.

- Giedroc, D. P. & Cornish, P. V. Frameshifting RNA pseudoknots: structure and mechanism. *Virus Res.* **139**, 193–208 (2009).
- Tsuchihashi, Z. Translational frameshifting in the *Escherichia coli* dnaX gene in vitro. *Nucleic Acids Res.* **19**, 2457–2462 (1991).
- Nackley, A. G. *et al.* Human catechol-O-methyltransferase haplotypes modulate protein expression by altering mRNA secondary structure. *Science* **314**, 1930–1933 (2006).
- Duan, J. B. *et al.* Synonymous mutations in the human dopamine receptor D2 (DRD2) affect mRNA stability and synthesis of the receptor. *Hum. Mol. Genet.* **12**, 205–216 (2003).
- Young, J. C. & Andrews, D. W. The signal recognition particle receptor alpha subunit assembles co-translationally on the endoplasmic reticulum membrane during an mRNA-encoded translation pause in vitro. *EMBO J.* **15**, 172–181 (1996).
- Watts, J. M. *et al.* Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature* **460**, 711–716 (2009).
- Takyar, S., Hickerson, R. P. & Noller, H. F. mRNA helicase activity of the ribosome. *Cell* **120**, 49–58 (2005).
- Wen, J. D. *et al.* Following translation by single ribosomes one codon at a time. *Nature* **452**, 598–603 (2008).
- Yusupova, G. Z., Yusupov, M. M., Cate, J. H. D. & Noller, H. F. The path of messenger RNA through the ribosome. *Cell* **106**, 233–241 (2001).
- Keller, D. & Bustamante, C. The mechanochemistry of molecular motors. *Biophys. J.* **78**, 541–556 (2000).
- Patel, S. S. & Donmez, I. Mechanisms of helicases. *J. Biol. Chem.* **281**, 18265–18268 (2006).
- Johnson, D. S., Bai, L., Smith, B. Y., Patel, S. S. & Wang, M. D. Single-molecule studies reveal dynamics of DNA unwinding by the ring-shaped T7 helicase. *Cell* **129**, 1299–1309 (2007).
- Lionnet, T., Spiering, M. M., Benkovic, S. J., Bensimon, D. & Croquette, V. Real-time observation of bacteriophage T4 gp41 helicase reveals an unwinding mechanism. *Proc. Natl Acad. Sci. USA* **104**, 19790–19795 (2007).
- Manosas, M., Xi, X. G., Bensimon, D. & Croquette, V. Active and passive mechanisms of helicases. *Nucleic Acids Res.* **38**, 5518–5526 (2010).
- Kim, S., Schroeder, C. M. & Xie, X. S. Single-molecule study of DNA polymerization activity of HIV-1 reverse transcriptase on DNA templates. *J. Mol. Biol.* **395**, 995–1006 (2010).
- Betterton, M. D. & Julicher, F. Opening of nucleic-acid double strands by helicases: active versus passive opening. *Phys. Rev. E* **71**, 011904 (2005).
- Tinoco, I. & Wen, J. D. Simulation and analysis of single-ribosome translation. *Phys. Biol.* **6**, 025006 (2009).
- Rodnina, M. V. *et al.* GTPase mechanisms and functions of translation factors on the ribosome. *Biol. Chem.* **381**, 377–387 (2000).
- Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**, 3406–3415 (2003).
- Tinoco, I. & Bustamante, C. The effect of force on thermodynamics and kinetics of single molecule reactions. *Biophys. Chem.* **101–102**, 513–533 (2002).
- Lohman, T. M., Tornko, E. J. & Wu, C. G. Non-hexameric DNA helicases and translocases: mechanisms and regulation. *Nature Rev. Mol. Cell Biol.* **9**, 391–401 (2008).
- Fischer, N., Konevega, A. L., Wintermeyer, W., Rodnina, M. V. & Stark, H. Ribosome dynamics and tRNA movement by time-resolved electron cryomicroscopy. *Nature* **466**, 329–333 (2010).
- Moazed, D. & Noller, H. F. Intermediate states in the movement of transfer RNA in the ribosome. *Nature* **342**, 142–148 (1989).
- Frank, J. & Agrawal, R. K. A ratchet-like inter-subunit reorganization of the ribosome during translocation. *Nature* **406**, 318–322 (2000).
- Schuwirth, B. S. *et al.* Structures of the bacterial ribosome at 3.5 Å resolution. *Science* **310**, 827–834 (2005).
- Valle, M. *et al.* Locking and unlocking of ribosomal motions. *Cell* **114**, 123–134 (2003).
- Peske, F., Matassova, N. B., Savelsbergh, A., Rodnina, M. V. & Wintermeyer, W. Conformationally restricted elongation factor G retains GTPase activity but is inactive in translocation on the ribosome. *Mol. Cell* **6**, 501–505 (2000).
- Chen, G., Chang, K. Y., Chou, M. Y., Bustamante, C. & Tinoco, I. Triplex structures in an RNA pseudoknot enhance mechanical stability and increase efficiency of 1-ribosomal frameshifting. *Proc. Natl Acad. Sci. USA* **106**, 12706–12711 (2009).
- Uemura, S. *et al.* Peptide bond formation destabilizes Shine–Dalgarno interaction on the ribosome. *Nature* **446**, 454–457 (2007).
- Cukras, A. R., Southworth, D. R., Brunelle, J. L., Culver, G. M. & Green, R. Ribosomal proteins S12 and S13 function as control elements for translocation of the mRNA: tRNA complex. *Mol. Cell* **12**, 321–328 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank members of the Tinoco and Bustamante labs for helpful discussions, especially S. B. Smith for his help with the optical tweezers, J. Moffitt for advice on data analysis and C. Kaiser for suggestions. Our work was supported by grants from the National Institutes of Health (to I.T., C.B. and H.F.N.) and the Human Frontiers Science Program (to I.T. and H.F.N.).

Author Contributions X.Q. and J.-D.W. conducted the experiments and performed the analysis; X.Q., J.-D.W. and L.L. prepared and provided experimental materials; and all authors helped to write the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to I.T. (intinoco@lbl.gov) or C.B. (carlos@alice.berkeley.edu).

CORRIGENDUM

doi:10.1038/nature10133

Sensitivity to antitubulin chemotherapeutics is regulated by MCL1 and FBW7

Ingrid E. Wertz, Saritha Kusam, Cynthia Lam, Toru Okamoto, Wendy Sandoval, Daniel J. Anderson, Elizabeth Helgason, James A. Ernst, Mike Eby, Jinfeng Liu, Lisa D. Belmont, Joshua S. Kaminker, Karen M. O'Rourke, Kanan Pujara, Pawan Bir Kohli, Adam R. Johnson, Mark L. Chiu, Jennie R. Lill, Peter K. Jackson, Wayne J. Fairbrother, Somasekar Seshagiri, Mary J. C. Ludlam, Kevin G. Leong, Erin C. Dueber, Heather Maecker, David C. S. Huang & Vishva M. Dixit

Nature **471**, 110–114 (2011)

During resubmission of this work, another paper linking MCL1 degradation and CDK1-mediated T92 phosphorylation in mitotic arrest was published¹. The studies use distinct approaches but come to similar conclusions.

1. Harley, M. E. Allan, L. A., Sanderson, H. S. & Clarke, P. R. Phosphorylation of Mcl-1 by CDK1–cyclin B1 initiates its Cdc20-dependent destruction during mitotic arrest. *EMBO J.* **29**, 2407–2420 (2010).

TECHNOLOGY FEATURE

INSIDE THE MINDS OF MICE AND MEN

Monitoring technologies and genetic engineering are producing a growing array of animal models for psychiatric disorders, but researchers are still learning how best to use them.

SAGE LABS



There is no such thing as an autistic rat, but researchers can study human psychiatric conditions by analysing the behaviour of rodents.

BY MONYA BAKER

Anyone familiar with Rett syndrome will recognize the symptoms. Mouse models — animals that carry genetic mutations similar to those that cause the condition in humans — wring their paws, walk awkwardly and learn poorly. Other human brain disorders have animal models, too. Mice with extra copies of the genome regions duplicated in Down's syndrome show motor problems and learning deficits. A developmental

disorder known as fragile X syndrome arises when humans lack a working copy of the gene *FMR1*; mice without the gene show learning deficits and hyperactivity similar to the symptoms of the human disorder.

But those are conditions with discrete, recognized causes. Other neurocognitive disorders, such as autism, depression and schizophrenia, have multiple and often mysterious causes, so mimicking them is more complicated. Studies have implicated dozens of genetic variants in producing the disorders, and environmental

factors from traumatic life experiences to *in utero* conditions also contribute. Even when researchers have decided which genes or factors to study, it is not always clear how to assess animal models: how can a researcher use a mouse to study diseases diagnosed by hallucinations or an inability to understand figurative language?

FROM BEHAVIOUR TO BIOLOGY

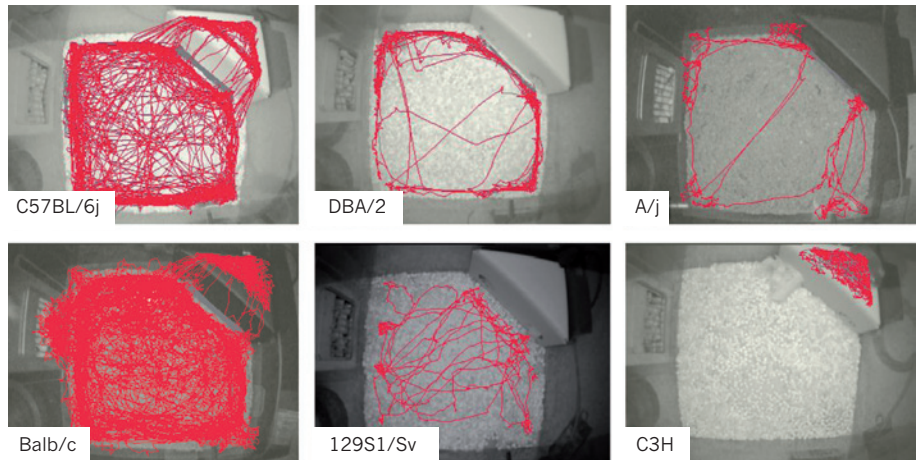
Craig Powell, a neuroscientist at the University of Texas Southwestern Medical Center in Dallas, says that the goal of tweaking genes is

usually to uncover disease mechanisms. “So you make a mouse with a mutation that you know causes autism in humans, and you see that it has behaviours that resemble autism,” he says. Powell’s next step is to look at slices of the mouse’s brain, to see how it differs from normal mice.

At that point, behaviour can help researchers to home in on what really matters. “We might find a hundred things wrong with the brain function, but only one or two cause changes in behaviour, so we want to fix things and see what changes the behaviour,” says Powell. In work that has become a touchstone for the field, Mark Bear, a neuroscientist at the Massachusetts Institute of Technology (MIT) in Cambridge, and his colleagues showed that reducing expression of a particular receptor in mice ameliorated the effects of a mutation that causes fragile X syndrome¹. Several physiological abnormalities were reversed, and engineered mice had fewer seizures and better memories.

In another example, Adrian Bird, a geneticist at the Wellcome Trust Centre for Cell Biology in Edinburgh, UK, and his colleagues reversed symptoms resembling Rett syndrome in mice². They crafted a version of the defective gene that could be restored to normal activity with a supplement to a mouse’s diet, but administered the supplement only after mice carrying the gene began to exhibit symptoms. Activating the gene at this point was expected to have little effect, but the mice showed marked improvement, raising hopes that recovery might also be possible in humans. A similar study³ into spinal muscular atrophy found that restoring a defective gene’s function four days after birth essentially eliminated signs of the disease; doing so ten days after birth had little effect. Such studies could help researchers to predict which patients are most likely to benefit in clinical trials.

Research into behaviour can also probe how genetic variants interact with each other and with environmental factors. In one study published this year⁴, wild-type males from five mouse strains were bred with females carrying a mutation that causes symptoms resembling



Video tracking of different mouse strains (labelled) reveals that some explore a new cage more than others.

autism and developmental disorders. Tests on the offspring showed that the symptomatic behaviours recurred in only some of the genetic backgrounds. In another study⁵, mice from a strain displaying a range of autism-relevant symptoms were reared by and with mice from a strain known for high sociability. The fostered mice showed no social deficits as adults, but other relevant symptoms, such as repetitive grooming, were not reduced. And when a mutant version of *DISC1*, the first gene to be implicated in schizophrenia, is present in mice whose mother’s immune system has been stressed during pregnancy, the offspring exhibit symptoms of affective disorders and autism⁶. Without the environmental stressors, they show symptoms of schizophrenia.

HIDDEN SYMPTOMS

Even the cleverest assays cannot capture some important aspects of human disease, such as the paranoid delusions common in schizophrenia. (In fact, because the models will always be imperfect, most behavioural researchers object to phrases such as ‘schizophrenic mice.’) Mikhail Pletnikov, a neurobehaviourologist at Johns Hopkins University in Baltimore, Maryland, is one of many scientists hoping to complement behavioural tests with more readily measured biomarkers. People with schizophrenia exhibit a wide range of behavioural symptoms, but their lateral ventricles — fluid-filled cavities on either side of the brain — tend to be larger than average, so Pletnikov is using brain scans to measure these structures in mice. It is not always necessary to see a behavioural change to probe a disease’s biology, he says. “With humility, you can use mice or rats or even worms.”

Unexpected behaviours have revealed unanticipated biology. Several years ago, Guoping Feng, a neuroscientist now at MIT, was trying to work out the function of various proteins found on either side of the synapses that connect neurons. Knocking out one such protein, SAPAP3, had no apparent effect on brain function: the mice walked and learned normally⁷. They did, however, seem to have something

wrong with their skin: open sores appeared on their faces. After tests showed nothing abnormal, video surveillance revealed that the mice groomed themselves excessively, literally rubbing through their fur. Further work showed abnormalities in a region of the brain linked to obsessive-compulsive disorder (OCD). Although SAPAP3 had not previously been implicated in the condition, drugs that eased OCD symptoms in humans reduced the grooming in mice.



“When in doubt, the experimenter needs to look at what the animal is doing.”

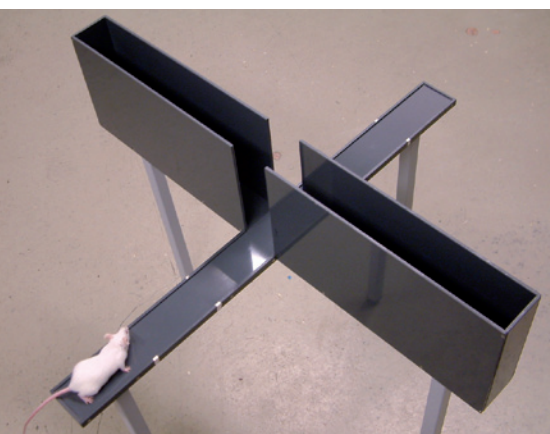
Douglas Wahlsten

Studies⁸ of proteins that interact with SAPAP3 revealed that they, too, had links to OCD and autism.

As human genetic studies reveal gene variants with increasingly smaller impacts on disease, there is increasing demand for new behavioural tests to assess them. Results of assays are highly variable, and they may not measure the most meaningful symptoms, says

Jeffrey Mogil, a neuroscientist at McGill University in Montreal, Canada. “We’ve made a lot of advances in making ever-fancier mice, but at the end of the day the question is, what’s your assay and what’s your measure and are they relevant?” he says. “The slow link in the chain, the messy link in the chain, has always been the behavioural assays.”

Current tests of animal behaviour are blunt tools. The Morris water-navigation task evaluates an animal’s cognitive ability by assessing how it learns to use spatial cues to swim to an underwater platform that it can’t see. Changes in the animal’s performance can be used to measure learning and memory. Tests for anxiety include the open-field test, which measures the time a mouse spends in enclosed spaces or along the edges of its cage;



Less-anxious mice spend longer in open spaces.

nervous animals avoid exposed areas. For both, psychiatric drugs effective in humans change the outcomes.

Such tests can be effective at screening new drugs that act by the same mechanisms as existing ones. But they are less useful for conditions for which no effective drugs exist, or for gaining insight into pathology. So researchers are trying to develop tests that capture more-specific components of human disorders.

“We talk more to the clinical researchers,” says Jacqueline Crawley, chief of behavioural neuroscience at the US National Institute of Mental Health in Bethesda, Maryland. “There are opportunities for us to sit down and say, ‘what do these diseases look like, what is their variability in the real world, and what do you consider the fundamental core symptoms?’”

Crawley’s own studies of children with autism inspired her to develop a mouse test for analogous behaviour. “You’ll see a group

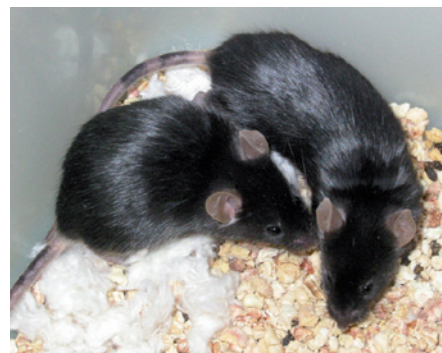


“An animal model may not be 100% translatable, but maybe 80% is good enough.”

Jacqueline Crawley

of children without autism playing together and the ones with autism being off to the side, playing with a train or a computer,” she recalls. So Crawley designed a task that would assess whether a mouse chose to spend time with a social partner or an inanimate object. The assay is now used in many laboratories.

Clinical tests have inspired other animal counterparts. Mogil and his colleagues produced the mouse-grimace scale for pain assessment⁹, based on a scale that used facial expressions to determine pain in infants and other humans incapable of speaking. Mogil believes his scale will prove a more reliable measure of chronic pain than commonly used assays such as the tail-flick test, which measures how quickly a mouse moves its tail out of a beam of light. It should also help researchers to design more-humane experiments. Other scientists have developed a mouse version¹⁰ of the Wisconsin card-sorting test, in which participants are presented with cards displaying, say, three red circles or four blue squares. Once humans recognize that rewards come for, say, matching cards by colour, the reward criteria are changed to matching by shape or number. The test is used to study disorders including autism and schizophrenia. The mouse version relies on scents such as cinnamon and garlic alongside textures such as gravel and cotton balls. Feng is currently evaluating the assay on



The colours of mice and their bedding can confuse video-tracking systems.

mouse models of autism.

Tim Bussey and Lisa Saksida, neuroscientists at the University of Cambridge, UK, have developed a mouse version of a touchscreen interface originally developed for humans and primates. It uses high-contrast images, tailored to mouse eyesight; and instead of pressing a lever or poking its nose into a hole in the wall as in most mouse-testing systems, the animal touches a screen with its nose or a paw. The technology, commercialized by Campden Instruments in Loughborough, UK, last year, could assess many cognitive abilities in rodents; one battery of tests assesses functions typically impaired in patients with schizophrenia, including visual perception,

D. WAHLSTEN/UNIV. NORTH CAROLINA GREENSBORO

working memory and pattern-learning¹¹. Bussey estimates that the system is now being used in more than 30 labs.

IRONING OUT THE KINKS

Confounding variables are the bane of behavioural testing, says Douglas Wahlsten, a neuroscientist at the University of North Carolina at Greensboro and the author of a handbook on the topic. For example, if a mouse seizes up with fear and stands still in the centre of an open chamber, the time it spends there could be misinterpreted as demonstrating reduced anxiety. And the walls of water-maze tanks are often so high above the water that mice cannot see much of the room, which makes it hard for them to use spatial cues to find the platform.

Genetic manipulation increases the scope for artefacts in the data, because tinkering with genes could alter how mice perform at tasks for reasons that have little to do with the parameters being tested (see 'Assessing the assays'). If learning assessments are based on an animals' ability to associate a sound with a mild electric shock, for example, researchers should make sure that the animals have normal hearing and sensitivity to pain. The biggest confounding variable may simply be moving the mouse from its cage to the area where tests are performed. "It's really rare that a small rodent would be lifted up by another animal and survive," says Laurence Tecott, a neurobiologist at the University of California, San Francisco. "We scare the hell out of the animal, then ask it if it's anxious and how it can learn," he says. "We do that routinely."

To reduce distress, researchers are working on ways not just to observe behaviour, but to do so without physically transporting animals first. In IntelliCage, an observation system from NewBehavior in Zurich, Switzerland, each animal is radiochipped. The system monitors when each animal drinks and eats, and how it performs at various stations in its enclosure.

Tecott, working with colleagues Evan Goulding and Katrin Schenk, has developed an inexpensive system that can monitor animals around the clock¹². A cage sits on a weight-detecting platform that measures an animal's location 50 times a second. Self-correcting informatics organize the animal's movements

Assessing the assays

Researchers evaluating animal models consider three kinds of validity.

Construct validity means that a test measures what it claims to. In animal models, that means that whatever causes symptoms in the animal is also what contributes to disease in humans. Such validity is relatively easy to achieve when a condition is caused by a single gene, but most are more complicated. Mikhail Pletnikov, a neurobiologist at Johns Hopkins University in Baltimore, Maryland, models schizophrenia by combining genes and environmental stressors. For complex disorders, he says, "we've passed that period where we manipulate one gene to try to understand the whole disease".

Face validity means that a test seems to measure what it needs to, for example that the symptoms in an animal model mirror those in a human. For heart rate or tumour growth, such measures may be straightforward, but for diseases assessed by behaviour, it is considerably more

complicated. BTBR mice, a strain used to study autism, avoid interacting with other mice and groom themselves excessively. When BTBR males are exposed to female urine, they do not vocalize and scent-mark as males from other strains do. These traits and others map well onto the diagnostic criteria for autism in humans, which include deficits in interaction and communication, along with repetitive behaviour.

Predictive validity is the extent to which a test predicts a future outcome. In an animal model, the animal should respond to drugs in a way that corresponds to human reactions. For example, antidepressants are sometime evaluated by their effects on the forced-swim test, which measures how long a mouse will try to climb out of a tank of water before giving up. For disorders such as autism, however, there are no effective drugs to serve as positive controls. Even when drugs do exist, the symptoms or mechanisms captured by a single behavioural assay are unlikely to capture everything that is important. **M.B.**

into 'bouts' of activity and can keep track of a mouse as it eats, defecates and moves its bedding. As part of the Mouse Phenome Project, an international collaboration to collect phenotypic data on mouse strains used in labs, Tecott is developing a lifestyle database for 16 strains. Even in preliminary results, strains can be distinguished by their distinct patterns of activity.

Tecott has also used his monitoring system on two lines of mice genetically engineered for obesity: *ob/ob* mice, which lack the gene to make one of the hormones that regulates appetite; and *htr2c* mutant mice, which lack a receptor for the neurotransmitter serotonin¹². The mice act like couch potatoes and midnight snackers respectively, says Tecott. The *ob/ob* animals eat just slightly more than mice of

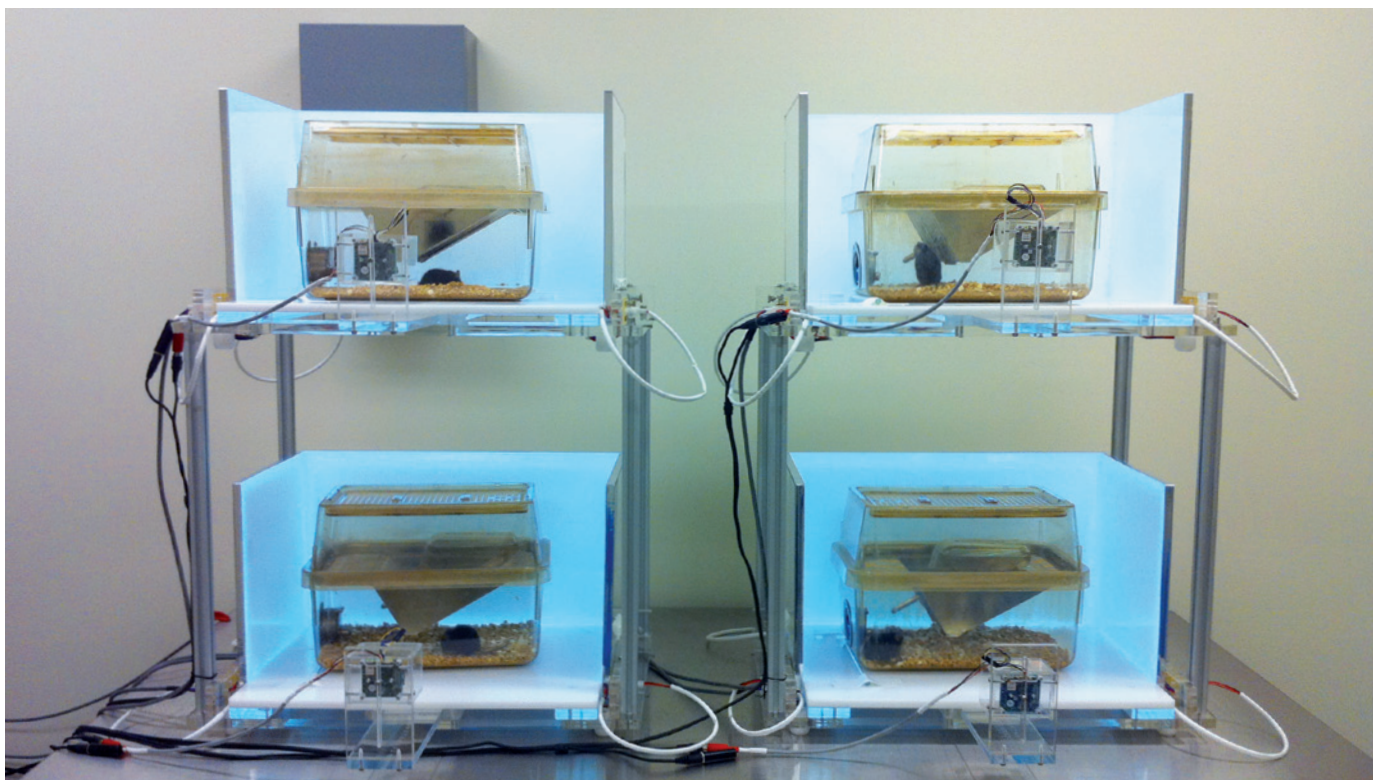
normal weight, but spend only about one-fifth as much time walking around their cages. The *htr2c* mutants have normal activity and feeding most of the time, but leave their burrows in the middle of their resting periods for a series of snacks. Without automated analysis, such insights into the behavioural components of obesity would be hard to detect.

More-expensive video-tracking systems, already widely used for many behavioural tests, can also be used to monitor animals in their home cages, automatically detecting and categorizing behaviours. As more molecular biologists want to monitor genetically engineered mice, demand for and applications of automated systems are increasing, says Lucas Noldus, chief executive of Noldus Information Technology in

SAGE LABS



Mice (left) are commonly used as disease models, but rats have more complex and sociable behaviour, so are better suited to modelling neurocognitive disorders.



Video-tracking systems analyse the behaviour of mice in their home cages. Software can capture specific activities, such as grooming and sniffing.

Wageningen, the Netherlands. Noldus describes one of his company's latest systems as "an instrumented home cage that can be configured in a variety of ways, from a very bare cage to very rich stimuli. Depending on the cage conformation you can perform all sorts of tests: an anxiety test with a light spot, or a memory test with automated pellet dispenser."

The monitoring component doesn't interfere with the animal, says Vikrant Kobla, vice-president of business development at Clever Sys in Reston, Virginia. "You're taking the same cage and putting a camera in front of it." His company's systems recognize more than two dozen behaviours, including head bobbing, grooming and standing on hind legs, and the repertoire is expanding. "We have so many modules we've developed that we can adapt it to deal with new behaviours," says Kobla.

CONSTANT VIGILANCE

"If you want a rich detail of measure from many kinds of tests, you really need to use video tracking," says Wahlsten. Nonetheless, he says, accuracy cannot be taken for granted, even for standard tests of animal behaviour. Tracking two animals at once is particularly difficult. It is not uncommon for software to confuse an animal's nose with its tail, for example. Bad lighting wrecks many experiments, particularly if a mouse is moving from a lighter area to a darker one, and different artefacts occur with white, brown, black and patchy mice. Infrared backlighting vastly reduces these problems and is commercially available, but is rarely used, owing to both lack of awareness and the cost

of converting existing equipment. But even the best systems require considerable effort to avoid false readings, warns Powell. "It's hard to sum up even in a book what the pitfalls are," he says. "Don't just take the Excel spreadsheet at the end and analyse it. Watch what's going on during the experiment."

Even when behaviours can be observed accurately, a mouse's activity may not be robust or subtle enough to reflect the effects of tweaking a gene or the environment. Rats show more complex behaviours; for example, littermates wrestle with each other, a behaviour that is considered social play. The most established behavioural tests were designed for rats, so many researchers are interested in modelling disease using genetically modified rats. Rats lacking genes implicated in schizophrenia, Parkinson's disease and autism are among the very first strains being produced by Sigma-Aldrich in Saint Louis, Missouri, which began offering a suite of ready-made knockout rats earlier this year. Sigma and other companies also take on custom projects to genetically engineer rats.

Richard Paylor, an autism researcher at Baylor College of Medicine in Houston, Texas, has just begun testing on knockout rats, investigating behaviours such as how the rats interact and vocalize in social situations, and how they respond to social odours. He hopes to report results by the end of the year. It is too early to say anything definitive, he says, but rats should allow finer behavioural assessments than mice. In particular, it may be possible to use them to quantify the effects of potential treatments for social and communication disorders,

something that has proved particularly difficult with mice. Rats' larger size also makes it easier to take electrophysiological recordings, brain images and tissue samples.

Paylor predicts that labs working with mice will find it difficult to switch to rats, which are expensive to buy and maintain, needing more space and different equipment. "We will be sort of a test case for labs that may want to study both mice and rats," says Shannon Hamilton, a postdoc in Paylor's laboratory.

But whether testing rats or mice, says Crawley, researchers must remember that the goal of an animal model is not perfection but utility. "An animal model may not be 100% translatable, but maybe 80% is good enough to test for possible treatments." ■

Monya Baker is technology editor for *Nature* and *Nature Methods*.

1. Krueger, D. D. & Bear, M. F. *Annu. Rev. Med.* **62**, 411–429 (2011).
2. Guy, J., Gan, J., Selfridge, J., Cobb, S. & Bird, A. *Science* **315**, 1143–1147 (2007).
3. Lutz, C. M. et al. *J. Clin. Invest.* (in the press).
4. Spencer, C. M. et al. *Autism Res.* **4**, 40–56 (2011).
5. Yang, M., Perry, K., Weber, M. D., Katz, A. M. & Crawley, J. N. *Autism Res.* **4**, 17–27 (2011).
6. Abazyan, B. et al. *Biol. Psychiatry* **68**, 1172–1181 (2010).
7. Welch, J. M. et al. *Nature* **448**, 894–900 (2007).
8. Peça, J. et al. *Nature* **472**, 437–442 (2011).
9. Langford, D. J. et al. *Nature Meth.* **7**, 447–449 (2010).
10. Bissonette, G. B. et al. *J. Neurosci.* **28**, 11124–11130 (2008).
11. Bussey, T. J. et al. *Neuropharmacology* doi:10.1016/j.neuropharm.2011.04.011 (2011).
12. Goulding, E. H. et al. *Proc. Natl Acad. Sci. USA* **105**, 20575–20582 (2008).

CAREERS

EDUCATION US needs to improve science literacy to prepare workforce **p.130**

TURNING POINT Biochemist's high-risk research direction pays off **p.131**

NATUREJOBS For the latest career listings and advice www.naturejobs.com



COLUMN

Turbocharge your writing today

Before you can tackle the overwhelming task of huge writing projects, you must first put aside some widely held myths, say **Maria Gardiner** and **Hugh Kearns**.

As a graduate student, you might find yourself well on the way with your education and 'ABD' (all but dissertation). Day after day, you tell yourself that you really, really intend to start writing your paper. After all, you've collected all the data, analysed them many times and entered them into tables.

But then you start thinking that maybe you need just a few more data. Perhaps, too, you should try a different analysis technique. And what if the tables you used aren't the right ones, or need to be formatted differently?

Many of the thousands of researchers we have worked with are constantly being tripped up by finicky, niggling details that keep them from writing up their research. Every day, they mean to start, but every day, something gets in their way or seems more important — and this can go on for years. Some very common obstacles get in the way of high-quality, high-quantity scholarly writing, but powerful, evidence-based techniques can help researchers to overcome repetitive and unhelpful habits and get moving (see 'How to get out of a dissertation-writing rut').

WRITING MYTHS

The biggest impediments to scholarly writing are long-held myths that seem to get passed down through the academic ranks like precious but unhelpful ancient wisdom. The first is the Readiness Myth — "I should write when I feel ready, and I don't feel ready yet". The secret to high output is that you have to write before you feel ready, because you might never reach that point. Researchers read endlessly and conduct countless experiments in the belief that it will eventually make them feel ready to write — we call these habits readitis and experimentitis. But ironically, all that reading and experimenting often makes them less likely to write, and more confused. So the first way to speed up your writing is to stop waiting, stop reading and experimenting, and start writing. You won't feel ready, but you have to do it anyway.

This brings us to the second myth, the Clarity Myth — "I should get it all clear in my head first, and then write it down". This isn't how writing works in practice. You have probably had the experience in which you were sure about how a paper would go until you started to write it. Then you discovered that there were inconsistencies, or it didn't flow well or the links didn't make sense. This tells you that it wasn't all that coherent in your head, after all. In fact, writing clarifies your thinking. Writing is not recording — you don't just take ►

EDUCATION

Better teaching needed

The United States must boost the number of people pursuing degrees and careers in science, technology, engineering and maths (STEM), says a 23 June report from the National Academies. The nation should foster better education in schools, said the report, *Successful K–12 STEM Education: Identifying Effective Approaches in Science, Technology, Engineering, and Mathematics*. The authors also recommend improving STEM literacy to fill STEM-related jobs that do not require advanced degrees, such as science teacher or energy technician. The US Bureau of Labor Statistics says that only 4 of the 16 STEM-related jobs with the largest projected growth by 2018 need an advanced degree.

JOB-HUNTING TOOLS

Inside information

An online forum aims to give job seekers inside information about employers. CareerBliss (www.careerbliss.com/company-questions) in Irvine, California, matches applicants with current employees who can answer queries. The forum has respondents for about 500 companies, universities and organizations in the United States including biopharmaceutical firms such as Pfizer and Genentech, says spokeswoman Alia Henson. Questions can be on any topic, including research funding or grant opportunities.

FAMILIES

Women want flexibility

Female early-career researchers with newborn babies are most likely to want to keep their jobs if their employers provide security and flexibility, including the right to leave work to care for an ill child, a study finds. Published on 23 May in the bi-monthly *Journal of Applied Psychology* (D. S. Carlson *et al.* *J. Appl. Psychol.* doi:10.1037/a0023964; 2011), the study reports better job retention for new mothers who stay physically and mentally healthy as a result of accommodations. Lead author Dawn Carlson, a professor of management at Baylor University in Waco, Texas, says that scientists should check how a prospective employer handles the needs of families before accepting an offer. To retain female staff, universities should allow maximum flexibility. “Whether extending the tenure clock or some other measure, the organization has to figure out a way to support these people if they want to reduce turnover,” says Carlson.

► a photocopy of what is in your head and put it on the page. It is a far more creative and interactive process. As you write, you develop your thoughts. Writing is, in fact, rigorous thinking. So the second way to turbocharge your writing and improve its quality is to get the words down on the page — no matter how bad you think they look or sound at first.

SNACK WRITING

Once researchers get beyond the myths that stop them writing, they often declare that they can't possibly write anything eloquent, insightful or clever unless they have a whole day or week to do it in. And because they don't have that amount of time, they conclude that there is no point in starting. We call this ‘binge writing’. Binge writing isn't inherently wrong; it's just that, for busy people, it can greatly reduce the amount of writing they do. The alternative is ‘snack writing’. This means short — but regular — writing sessions. We suggest about 1–2 hours a day for graduate students who are writing a dissertation, and about 45–90 minutes a day for researchers trying to increase their publication output.

Many researchers tell us that they couldn't possibly get anything useful written in that amount of time. The good news is that studies (which we have replicated many times in practice) show that academics who write for 30 minutes a day produce, on average, more peer-reviewed publications than academics who write for big blocks of time. But the ‘snacks’ have to be regular — 45 minutes once a week doesn't work, but 45 minutes a day 5 days a week does wonders. When possible, try snack writing first thing in the morning. Our experience suggests that this increases the

chances of success by minimizing distractions and ensuring that you have sufficient energy to write clever things. However, for snack writing to lead to really high-quality results, you also need to write in a very specific way.

WHAT IS WRITING?

Before we tell you what writing is, we should tell you what it isn't, at least for the purposes of snack writing.

Writing isn't editing: you should not spend your brief snack-writing time trying to find the perfect word or getting your grammar right. Writing isn't reading journal articles for research: write first and read afterwards, so that your writing shows you what you need to read. Writing isn't referencing: when you make that killer argument and want to reference Smith and Brown (2006; or maybe it was 2007?), don't stop and look it up. Write “Smith & Brown (200??)” and keep going. You can look up the reference later. Furthermore, writing is not formatting, literature searching, photocopying, e-mailing or nosing around on Facebook. Writing — at least for your snack-writing sessions — means putting new words on the page or substantially rewriting existing words.

So, you might ask, when do you do all the editing, reading and other associated tasks? The answer is, any time in the other 23 hours and 15 minutes of the day — just not during your snack-writing time.

So stop waiting to feel ready. Get started with some short and regular writing snacks. What you write won't be perfect at first, but you will be on your way to becoming a prolific academic writer. ■

Maria Gardiner and Hugh Kearns lecture and research in psychology at Flinders University in Adelaide, Australia, and run workshops for graduate students and advisers (see ithinkwell.com.au).

TOP TIPS

How to get out of a dissertation-writing rut

- Write before you feel ready — because you might never feel ready. It's amazing how people magically feel ready when there is an imminent deadline.
- Don't wait to have a clear picture of the paper. As you start putting down your ideas, you may actually clarify them.
- Snack write — work in short, frequent bursts instead of waiting to sit down for big blocks of time. Those blocks hardly ever come, and when they do, they don't usually get used very productively.
- Set specific times in your schedule for writing — don't leave it to chance, because

chances are it won't happen.

- Writing means putting new words on the page or substantially rewriting old words. It does not mean editing, reading, referencing or formatting — and it definitely does not mean composing e-mails.
- If you refrain from writing because you worry that what you write won't be good enough, try noting the adage that to write well, you first have to write.
- To really increase the quality and quantity of your writing, get feedback from mentors and colleagues — it can be painful, but it works. **M.G. and H.K.**

EDUCATION

Better teaching needed

The United States must boost the number of people pursuing degrees and careers in science, technology, engineering and maths (STEM), says a 23 June report from the National Academies. The nation should foster better education in schools, said the report, *Successful K–12 STEM Education: Identifying Effective Approaches in Science, Technology, Engineering, and Mathematics*. The authors also recommend improving STEM literacy to fill STEM-related jobs that do not require advanced degrees, such as science teacher or energy technician. The US Bureau of Labor Statistics says that only 4 of the 16 STEM-related jobs with the largest projected growth by 2018 need an advanced degree.

JOB-HUNTING TOOLS

Inside information

An online forum aims to give job seekers inside information about employers. CareerBliss (www.careerbliss.com/company-questions) in Irvine, California, matches applicants with current employees who can answer queries. The forum has respondents for about 500 companies, universities and organizations in the United States including biopharmaceutical firms such as Pfizer and Genentech, says spokeswoman Alia Henson. Questions can be on any topic, including research funding or grant opportunities.

FAMILIES

Women want flexibility

Female early-career researchers with newborn babies are most likely to want to keep their jobs if their employers provide security and flexibility, including the right to leave work to care for an ill child, a study finds. Published on 23 May in the bi-monthly *Journal of Applied Psychology* (D. S. Carlson *et al.* *J. Appl. Psychol.* doi:10.1037/a0023964; 2011), the study reports better job retention for new mothers who stay physically and mentally healthy as a result of accommodations. Lead author Dawn Carlson, a professor of management at Baylor University in Waco, Texas, says that scientists should check how a prospective employer handles the needs of families before accepting an offer. To retain female staff, universities should allow maximum flexibility. “Whether extending the tenure clock or some other measure, the organization has to figure out a way to support these people if they want to reduce turnover,” says Carlson.

► a photocopy of what is in your head and put it on the page. It is a far more creative and interactive process. As you write, you develop your thoughts. Writing is, in fact, rigorous thinking. So the second way to turbocharge your writing and improve its quality is to get the words down on the page — no matter how bad you think they look or sound at first.

SNACK WRITING

Once researchers get beyond the myths that stop them writing, they often declare that they can't possibly write anything eloquent, insightful or clever unless they have a whole day or week to do it in. And because they don't have that amount of time, they conclude that there is no point in starting. We call this ‘binge writing’. Binge writing isn't inherently wrong; it's just that, for busy people, it can greatly reduce the amount of writing they do. The alternative is ‘snack writing’. This means short — but regular — writing sessions. We suggest about 1–2 hours a day for graduate students who are writing a dissertation, and about 45–90 minutes a day for researchers trying to increase their publication output.

Many researchers tell us that they couldn't possibly get anything useful written in that amount of time. The good news is that studies (which we have replicated many times in practice) show that academics who write for 30 minutes a day produce, on average, more peer-reviewed publications than academics who write for big blocks of time. But the ‘snacks’ have to be regular — 45 minutes once a week doesn't work, but 45 minutes a day 5 days a week does wonders. When possible, try snack writing first thing in the morning. Our experience suggests that this increases the

chances of success by minimizing distractions and ensuring that you have sufficient energy to write clever things. However, for snack writing to lead to really high-quality results, you also need to write in a very specific way.

WHAT IS WRITING?

Before we tell you what writing is, we should tell you what it isn't, at least for the purposes of snack writing.

Writing isn't editing: you should not spend your brief snack-writing time trying to find the perfect word or getting your grammar right. Writing isn't reading journal articles for research: write first and read afterwards, so that your writing shows you what you need to read. Writing isn't referencing: when you make that killer argument and want to reference Smith and Brown (2006; or maybe it was 2007?), don't stop and look it up. Write “Smith & Brown (200??)” and keep going. You can look up the reference later. Furthermore, writing is not formatting, literature searching, photocopying, e-mailing or nosing around on Facebook. Writing — at least for your snack-writing sessions — means putting new words on the page or substantially rewriting existing words.

So, you might ask, when do you do all the editing, reading and other associated tasks? The answer is, any time in the other 23 hours and 15 minutes of the day — just not during your snack-writing time.

So stop waiting to feel ready. Get started with some short and regular writing snacks. What you write won't be perfect at first, but you will be on your way to becoming a prolific academic writer. ■

Maria Gardiner and Hugh Kearns lecture and research in psychology at Flinders University in Adelaide, Australia, and run workshops for graduate students and advisers (see ithinkwell.com.au).

TOP TIPS

How to get out of a dissertation-writing rut

- Write before you feel ready — because you might never feel ready. It's amazing how people magically feel ready when there is an imminent deadline.
- Don't wait to have a clear picture of the paper. As you start putting down your ideas, you may actually clarify them.
- Snack write — work in short, frequent bursts instead of waiting to sit down for big blocks of time. Those blocks hardly ever come, and when they do, they don't usually get used very productively.
- Set specific times in your schedule for writing — don't leave it to chance, because

chances are it won't happen.

- Writing means putting new words on the page or substantially rewriting old words. It does not mean editing, reading, referencing or formatting — and it definitely does not mean composing e-mails.
- If you refrain from writing because you worry that what you write won't be good enough, try noting the adage that to write well, you first have to write.
- To really increase the quality and quantity of your writing, get feedback from mentors and colleagues — it can be painful, but it works. **M.G. and H.K.**

TURNING POINT

Job Dekker

Job Dekker won the 2011 Young Investigator Award from the American Society for Biochemistry and Molecular Biology in Bethesda, Maryland, for his efforts to develop techniques to probe the three-dimensional structure of chromosomes. Dekker, a biochemist at the University of Massachusetts Medical School in Worcester, accepted the honour on 12 April.

What was the most important factor in your career development?

Where I did my postdoc. I did my PhD in biochemistry at Utrecht University in the Netherlands, but I had always been interested in how chromosomes interact and exchange information. I followed the literature on this issue, and when it was time to set up a postdoc, I contacted Nancy Kleckner, a molecular biologist at Harvard University in Cambridge, Massachusetts. She was studying how chromosome structure dictates where and when recombination occurs, which dovetailed with my interest in mapping chromosome structure.

Describe the roots of your technique.

I developed a biochemical technique for determining how DNA segments are linked to one another and interact. The result is akin to a 'molecular microscope' to detect physical interactions between chromosomes, and it eventually became the chromosome conformation capture (3C) method. 3C has helped my team and others to study how chromosomal activities influence, for example, gene expression. Nancy was supportive of this wild idea and was willing to keep me in the lab for as long as it took to get it to work.

How did you approach this high-risk project?

I gave myself 12 months to collect experimental evidence that it would work. A year isn't long, but a postdoc is the right time to do a risky project. If it works out, it's a great starting point for the rest of your career. If you stick to safe projects, you carve out a less distinct niche. I started in January of that year and got the first hint that it would work in November. But I immediately put it aside.

Why?

During my postdoc, I was also doing a 'safe' project — studying how proteins mediate interactions between chromosomes. More labs were entering the competition, so I had to pause the high-risk project to focus on the safe one. But I got scooped on that, so I went back to the microscope project, which took



two more years. I published the 3C paper four years into my postdoc, which is a long time to go without papers. But it's an important one.

Did you worry about cultivating a reputation as a technologist rather than as a scientist?

Yes, and I discussed the issue with my department chair. But quite a few career opportunities came my way as a result of the 3C work. I forged collaborations with Eric Lander, founder of the Broad Institute genomic-medicine research centre in Cambridge, Massachusetts, and his sequencing group, and we developed the technology further, using it to visualize whole genomes, and ultimately combining it with next-generation sequencing to create a high-throughput version of 3C. And we published all the intermediate technologies, so I have a solid publication record.

What are the consequences of sharing your technological developments?

It's both beneficial and harmful. We train people from other labs and give them our protocols to help promote advances. The benefit is that I have a whole network of connections, and the technology has rapidly taken hold. As a result, we've got a lot of competition. We've been scooped several times because the technology is now so easy to gain access to. But that's normal in science.

You've just achieved tenure. Is your career in transition?

My research direction is changing. After spending 13 years on this technology, we can now look at whole chromosomes to address real questions, such as how chromosome structure affects disease formation. I don't think I'll embark on a new round of technology development, but you never know. ■

INTERVIEW BY VIRGINIA GEWIN

AN EVANESCENT BOOK

The gift of knowledge.

BY TITO URETA

A recension of the Book of Sand

ANONYMOUS, DESCRIBED BY JORGE LUIS BORGES, ALIANZA EDITORIAL BUENOS AIRES 1977.

Quantum Foam Holographic Publishers, London, 2031. Quantumholographic edition. ∞ pp.

From about 2030 onwards there were rumours that a group of librarians had finally found the *Book of Sand*, which Jorge Luis Borges purposely lost (among maps and periodicals) in the cavernous depths of the National Library in Buenos Aires. As will be remembered, the book contains an infinite number of pages, none of them being the first and none being the last. Even though Borges did not say so explicitly, several scholars have attributed the unbounded vastness of this mythical book to the fact that it is the repository of all knowledge, in the form of an encyclopaedia with pages that divide and subdivide without limit. When Borges examined the book, the World-Wide Web did not exist but his description is quite clear. The *Book of Sand* is not the Web as it is truthful, non-profit, non-redundant and garbage-free.

The *Book of Sand* has indeed been found and it is now available for examination. It is certainly impossible to produce a facsimile or a digital version with conventional techniques, as the content of the book is infinite and the digitization would require a computer with infinite memory. Nevertheless, with a newly described technique whose scientific basis eludes me, known as Quantum Foam Holographic Scanning, it is now possible to produce a hologram that can be perused in the same way as if the original were in your hands.

Everything that Borges said about the *Book of Sand* is true. Like him, I tried to open the first or the last page, but no matter what I tried a few pages inevitably appeared between the cover and my fingers. If I opened it at random I always found two facing pages of entries, some of them with pictures, as in a conventional encyclopaedia.

Because of my professional bias I was delighted to find an entry labelled 'DNA'. It begins with a description of Miescher's experiments, followed by those of Griffiths (without mentioning the names of the scientists), but on reaching the bottom line and attempting to turn the page, I found myself in front of a page describing the adventures of a medieval gentleman named Durandarte. When I tried to go back to the previous page I found that it was no longer possible.

I wonder about those pages that I can no longer find. Of course the enormous discovery by Avery, the telling findings of Chargaff, the fecund model of Crick and Watson, and the Genome Project of the end of the last century must all be there, and, maybe, the base sequences of the 300,000 genomes now in the databases (from all creatures living and extinct?) and surely the current research on extragalactic DNA. I am sure that all that is displayed in the pages between DNA and Durandarte but the probability of opening one of them is

vanishingly small, and if I get close I shall perhaps find a passage about Donosti-
arra or the love potions of a scoundrel called Dulcamara. Perhaps — horror! — only very long and very boring series of purines and pyrimidines in a list like TTTGGTCAGATG-GCATTGGG... from an organism whose name appears in a page that I cannot visit.

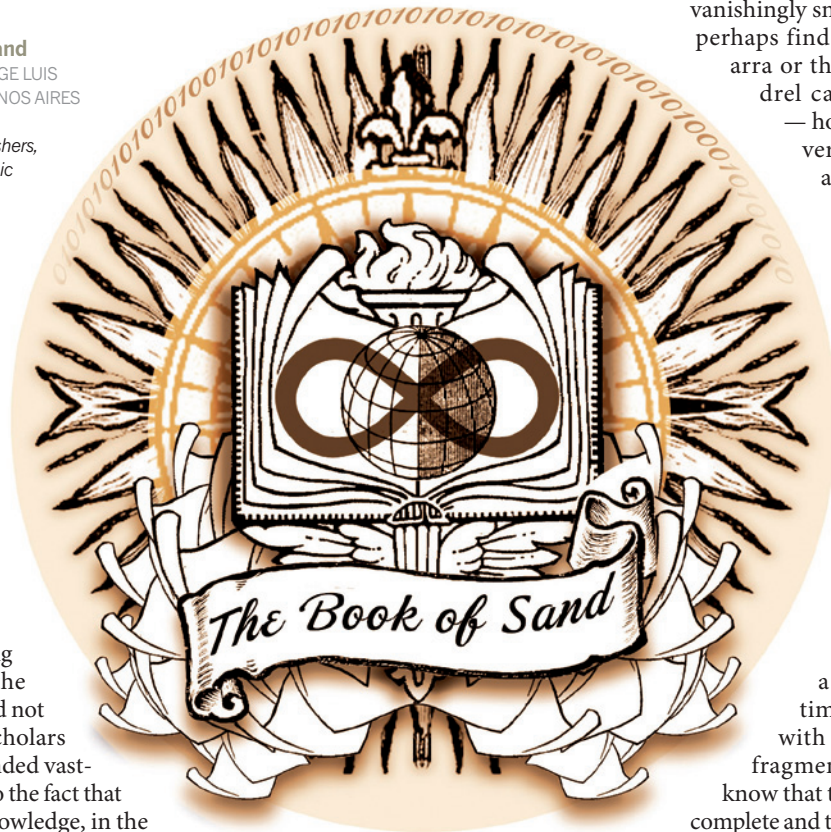
The *Book of Sand* is, then, a mirror of our world. We use all our senses trying to make sense of it. But the world, like the *Book of Sand*, allows us only to retrieve, *per speculum et in aenigmate*, mere bits that only rarely give us the certainties that we need.

Thus the *Book of Sand* is a complete, and at the same time a partial, encyclopaedia with pages that can offer only fragments. But it is infuriating to know that the encyclopaedia is, in fact, complete and that all the present and future knowledge is there, if only one could turn the pages in the proper sequence. Alas! the book is a sandbook and it is therefore impossible to do that.

But then, it might be better that way. If I could read the *Book of Sand* as an ordinary encyclopaedia, it would no longer be worthwhile to continue my ongoing research on the molecular dictionaries of the Nostratic language. If I could find in the *Book of Sand* the appropriate paragraphs, I would know today all about the molecular structure of the grammars stored in the brain and, therefore, all the fun of pursuing my investigations on that marvellous area would be gone for ever. It was indeed a wise man who said that the path contains the goal and that the goal cannot exist without the path.

So I have closed my hologram copy of the *Book of Sand* and put it on a nearby shelf. Who knows? From time to time I shall open it at random and maybe I shall find some nuggets that shed light on my questions. ■

Tito Ureta is a professor in the Departamento de Biología, Facultad de Ciencias, Universidad de Chile, Santiago de Chile.



► **NATURE.COM**
Follow Futures on
Facebook at:
go.nature.com/mtoodm

Coordination of DNA replication and histone modification by the Rik1–Dos2 complex

Fei Li¹, Rob Martienssen² & W. Zacheus Cande³

Histone modification marks have an important role in many chromatin processes^{1,2}. During DNA replication, both heterochromatin and euchromatin are disrupted ahead of the replication fork and are then reassembled into their original epigenetic states behind the fork^{3,4}. How histone marks are accurately inherited from generation to generation is still poorly understood. In fission yeast (*Schizosaccharomyces pombe*), RNA interference (RNAi)-mediated histone methylation is cell cycle regulated. Centromeric repeats are transiently transcribed in the S phase of the cell cycle and are processed into short interfering RNAs (siRNAs) by the complexes RITS (RNA-induced initiation of transcriptional gene silencing) and RDRC (RNA-directed RNA polymerase complex)^{5–7}. The small RNAs together with silencing factors—including Dos1 (also known as Clr8 and Raf1), Dos2 (also known as Clr7 and Raf2), Rik1 and Lid2—promote heterochromatic methylation of histone H3 at lysine 9 (H3K9) by a histone methyltransferase, Clr4 (refs 8–13). The methylation of H3K9 provides a binding site for Swi6, a structural and functional homologue of metazoan heterochromatin protein 1 (HP1)¹⁴. Here we characterize a silencing complex in fission yeast that contains Dos2, Rik1, Mms19 and Cdc20 (the catalytic subunit of DNA polymerase-ε). This complex regulates RNA polymerase II (RNA Pol II) activity in heterochromatin and is required for DNA replication and heterochromatin assembly. Our findings provide a molecular link between DNA replication and histone methylation, shedding light on how epigenetic marks are transmitted during each cell cycle.

To explore the role of Dos2 in heterochromatin assembly, we sought to identify Dos2-associated proteins by using tandem affinity

purification (TAP). Mass spectrometry analysis of Dos2 purified by the TAP method uncovered two new interacting proteins, in addition to Rik1: Cdc20 (ref. 15); and a previously uncharacterized protein, SPAC1071.02 (Fig. 1a). SPAC1071.02 is highly conserved (Supplementary Fig. 1). Its homologue in budding yeast (*Saccharomyces cerevisiae*) is MMS19 (ref. 16), so we named the fission yeast protein Mms19. The interactions of Dos2 with Mms19 and Cdc20 were confirmed by co-immunoprecipitation experiments (Supplementary Figs 2 and 3).

Cdc20 is a conserved DNA polymerase-ε subunit with extensive homology to its counterparts in humans and budding yeast. Cdc20 regulates the elongation of the leading strand during DNA replication, shortly after initiation, and is essential for cell viability¹⁵. To test whether Cdc20 is required for heterochromatin silencing, we used a temperature-sensitive mutant allele, *cdc20-p7*. At 37 °C, mutant cells arrest in early S phase. We crossed these mutants into the *otr::ura4⁺* background and performed a silencing assay. We found that, at a non-restrictive temperature, 34 °C, the mutant cells grew poorly on control medium compared with wild-type (WT) cells (Fig. 1b), probably as a result of the replication abnormality. However, on medium lacking uracil, the mutants had more robust growth than WT cells. By contrast, on 5-fluoroorotic acid (5-FOA)-containing medium, the mutant cells had little growth, demonstrating that centromeric silencing was partially compromised. These results indicate that complete silencing of heterochromatin requires Cdc20.

In WT cells, heterochromatin transcripts are quickly processed by the RNAi machinery, but in RNAi-processing defective mutants, such as *dcr1-Δ*, these transcripts are readily detectable¹⁷. In *cdc20-p7* mutant

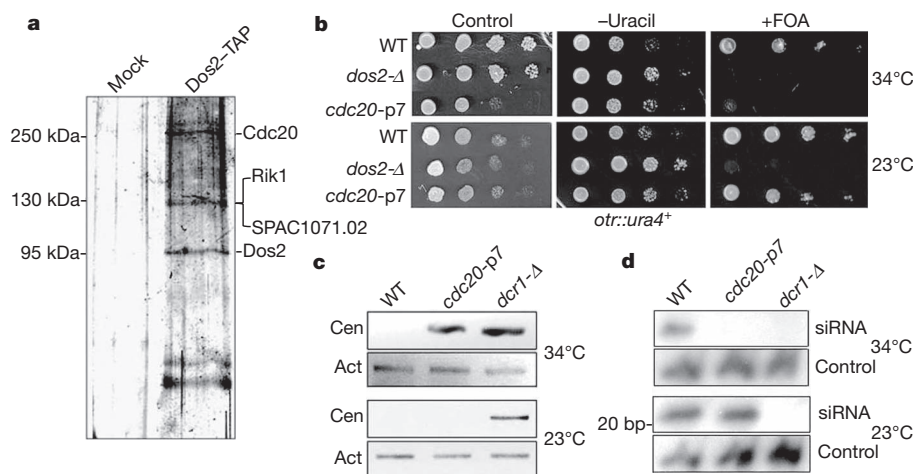


Figure 1 | Cdc20 is essential for transcriptional silencing and siRNA generation. **a**, Protein extracts from a Dos2–TAP strain or an untagged control strain (mock) were purified by the TAP method. Purified products were separated by SDS–PAGE and visualized by silver staining. kDa, kilodalton. **b**, Growth assay of serial dilutions (left to right) of strains carrying *ura4⁺* inserted at the pericentromeric *otr* region, grown on control medium, medium

lacking uracil (–uracil) or counter-selective 5-FOA medium (+FOA) incubated at 23 °C or 34 °C. **c**, Accumulation of centromeric transcripts analysed by RT–PCR in strains incubated at the indicated temperatures. Act, *act1⁺* control; Cen, centromeric transcripts. **d**, Analysis of siRNAs corresponding to centromeric repeats by northern blotting. Control, snoR69 as a loading control.

¹Department of Biology, New York University, New York, New York 10003, USA. ²Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA. ³Department of Molecular and Cell Biology, University of California, Berkeley, California 94720, USA.

cells incubated at 34 °C, similar to *dcrl-Δ* mutants, pericentromeric transcripts accumulated (Fig. 1c). We then examined the amount of small RNA in the *cdc20-p7* mutant by northern blotting. Because RNAi is temperature sensitive, the abundance of small RNAs in WT cells was considerably reduced at 34 °C but was still detectable⁶ (Fig. 1d). In the *cdc20-p7* mutant, however, siRNAs were completely absent (Fig. 1d), showing that Cdc20 promotes siRNA generation.

To further determine how heterochromatin structure is affected by disruption of Cdc20, we examined H3K9 methylation and Swi6 distribution in the *cdc20-p7* cells grown at 34 °C and 23 °C. Pericentromeric H3K9 methylation was significantly reduced at the higher temperature (Fig. 2a) and the association of Swi6 was also decreased (Supplementary Fig. 4), both of which are consistent with the heterochromatin defect shown by the silencing assay. We also assessed the delocalization of green fluorescent protein (GFP)–Swi6 fusion proteins to determine loss of heterochromatin, because the GFP–Swi6 pattern is unchanged in RNAi mutants¹⁸. We found that 53% of *cdc20-p7* cells at 34 °C, and more than 70% at 37 °C, had a diffuse GFP–Swi6 pattern, a defect similar to the *dos2-Δ* mutant (Supplementary Fig. 5 and Fig. 2b). WT cells incubated at the elevated temperatures did not show severe Swi6 delocalization (Fig. 2b), demonstrating that heterochromatin formation requires Cdc20. Because heterochromatin formation is mediated by both RNAi-dependent and RNAi-independent pathways¹⁹, the silencing abnormality in *cdc20-p7* mutants indicated that Cdc20 functions at an early stage of heterochromatin assembly.

Mms19, another Dos2-interacting factor, is a conserved protein that contains a HEAT repeat domain (Supplementary Fig. 1). Studies of its homologues in budding yeast and humans show that they function as regulators of the transcription factor TFIIH, participating in the initiation of RNA–Pol-II-mediated transcription^{16,20}. Interestingly, human *MMS19* is also required for chromosome segregation²¹. To

study the role of Mms19 in fission yeast, we first examined its distribution using a GFP-tagged version of Mms19 and found that the GFP signal was predominantly nuclear, consistent with its potential role as a transcriptional regulator (Supplementary Fig. 6). To elucidate the function of Mms19, we created an *mms19*-null (*mms19-Δ*) mutant. This mutant grew slower than the WT strain but was viable, indicating that *mms19*⁺ is not an essential gene. Similar to the budding yeast *mms19* mutant, the growth of *mms19-Δ* required methionine (Supplementary Fig. 7).

Because Mms19 homologues in other organisms associate with the TFIIH complex, we speculated that Mms19 might be involved in RNA–Pol-II-mediated transcription in heterochromatic regions. To address this possibility, we directly examined centromeric transcription by using PCR with reverse transcription (RT–PCR). Centromeric transcripts were abundant in siRNA-processing mutants, such as *dcrl-Δ*, but it was difficult to detect them in the WT strain¹⁷ (Supplementary Fig. 8). In contrast, these transcripts were not discernible in the *mms19-Δ* mutant by using RT–PCR, similarly to the WT strain (Supplementary Fig. 8), and their abundance was greatly reduced in a *dcrl-Δ mms19-Δ* double mutant (Fig. 3a). We reasoned that, as a result of the reduction in primary siRNA transcripts, centromeric siRNA levels might also be decreased. To test this, RNA extracted from the *mms19-Δ* mutant was probed for centromeric siRNA by northern blotting. These siRNAs were present but less abundant than in the WT strain (Fig. 3b). These data further demonstrate that Mms19 regulates centromeric transcription.

Coincident with heterochromatin expression, RNA Pol II is preferentially restricted to heterochromatin in S phase⁵. To further elucidate the role of Mms19 in this process, we investigated how Mms19 associates with heterochromatin during the cell cycle. After release from synchronization, cells carrying Mms19–TAP were collected at different

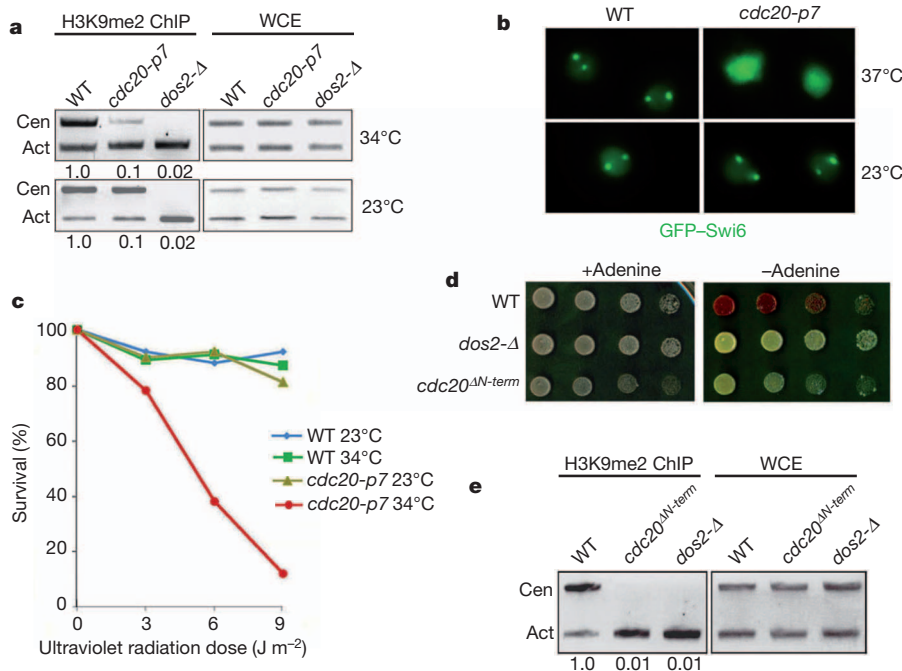


Figure 2 | Heterochromatin abnormality is coupled to DNA replication defects. **a**, Pericentromeric H3K9 methylation in *cdc20-p7* was significantly lost at 34 °C but not at 23 °C. A ChIP assay was performed with an antibody specific for dimethylated H3K9 (H3K9me2). DNA from precipitates was analysed by competitive PCR in which one set of primers amplified the centromeric *dh* repeat (Cen) and another amplified the control gene *act1*⁺ (Act). The relative fold enrichment (indicated below each lane) was calculated by comparing the ratios of heterochromatin signals to control signals in the ChIP and whole cell extract (WCE) fractions. The value for WT cells was set to 1.0. **b**, Fluorescent images of GFP–Swi6-carrying strains incubated at 23 °C or 37 °C. GFP–Swi6 localizes to heterochromatin regions in the form of two to

four green foci in WT cells in interphase. This localization was lost in more than 70% of *cdc20-p7* cells at 37 °C. **c**, Ultraviolet radiation sensitivity of the *cdc20-p7* mutant and the WT strain at 23 °C or 34 °C. **d**, Colony colour silencing assay of strains carrying *ade6*⁺ inserted at the pericentromeric *otr* region, grown at 23 °C on YES medium with or without further supplementation with adenine. Repression of *ade6*⁺ expression results in a red or pink colour; when transcriptional silencing does not occur, the colonies appear white. **e**, Analysis of pericentromeric H3K9 methylation in the indicated strains. A ChIP assay was carried out with an antibody specific for H3K9me2, followed by PCR, as in **a**. The relative fold enrichment of the methylation is indicated below each lane.

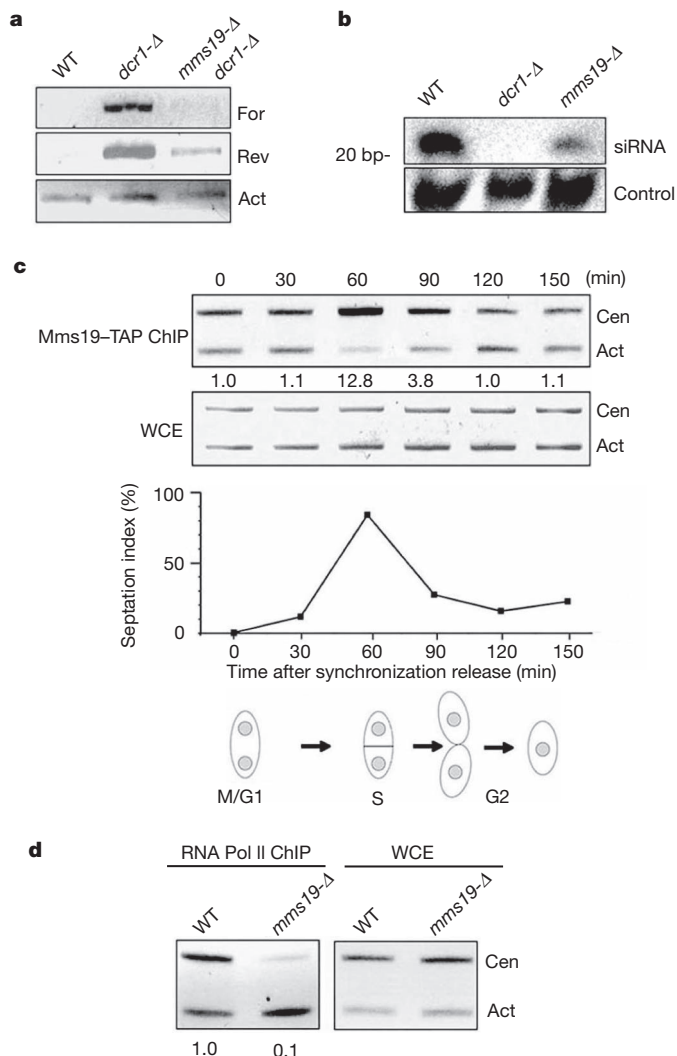


Figure 3 | Mms19 is required for RNA-Pol-II-mediated transcription of heterochromatin. **a**, Strand-specific RT-PCR analysis of the accumulation of transcripts from centromeric *dh* repeats. For, forward strand; Rev, reverse strand. **b**, Analysis of siRNA corresponding to centromeric *dg-dh* repeats by northern blotting. bp, base pairs; control, snoR69 as a loading control. **c**, Mms19 preferentially associates with heterochromatin during S phase. Protein extracts were prepared from synchronized cells carrying Mms19-TAP, which were then analysed at various time points after synchronization release by using ChIP with an antibody specific for TAP, followed by PCR as in Fig. 2a. The relative fold enrichment of Mms19 is indicated below each lane. Cell cycle progression was monitored by a septation index, which is also illustrated diagrammatically. **d**, ChIP analysis of RNA Pol II accumulation in pericentromeric heterochromatin in cells synchronized in S phase, followed by PCR, as in c. The relative fold enrichment of RNA Pol II is indicated below each lane.

stages of the cell cycle. We found that Mms19 preferentially associated with heterochromatin in S phase (Fig. 3c), concurrently with the enrichment of RNA Pol II in heterochromatin. We then investigated how Mms19 affects the RNA Pol II distribution in heterochromatin at this stage. Using chromatin immunoprecipitation (ChIP) with an antibody specific for RNA Pol II, we found that the RNA Pol II accumulation in S phase was reduced considerably in the *mms19-Δ* mutant (Fig. 3d). Furthermore, Mms19 was found to physically associate with RNA Pol II (Supplementary Fig. 9). Together, our results suggest that Mms19 is a transcriptional activator that is required for the RNA-Pol-II-mediated transcription of heterochromatin.

To gain further insight into how Cdc20 interacts with Dos2 and Mms19, we created a haemagglutinin (HA)-tagged version of the

mutant gene from the *cdc20-p7* strain. Co-immunoprecipitation experiments showed that Cdc20-P7-HA maintained its association with Dos2 and Mms19 at 23 °C; however, these interactions were lost at 34 °C, indicating that the point mutation reduces the interactions at the elevated temperature (Fig. 4a). We also investigated the association between Mms19 and heterochromatin by using ChIP in synchronized *cdc20-p7* cells released from metaphase. At 23 °C, there was a clear peak in Mms19 enrichment in the mutant during S phase, which occurred ~60 min after release from synchronization; however, this accumulation of Mms19 was not observed when the temperature was elevated to 34 °C, indicating that Cdc20 is required for the association of Mms19 with heterochromatin (Fig. 4b). A previous report showed that Dos2 and Rik1 start to accumulate in heterochromatin in S phase⁵. We therefore assessed whether Cdc20 affects the recruitment of these two silencing factors. Using ChIP with antibodies specific for TAP or a Myc tag, we found that Dos2-TAP and Rik1-Myc are enriched in S phase at 23 °C, consistent with the previous study; however, the association is diminished at 34 °C (Fig. 4c and Supplementary Fig. 10). These results indicate that Cdc20 is required for the recruitment of Dos2 and Rik1 to heterochromatin. Interestingly, heterochromatin silencing is also partly compromised in strains with mutations in two different DNA polymerase- α subunits^{22,23}. As DNA polymerase- α primase is required before elongation by DNA polymerase- ϵ (of which Cdc20 is a subunit), it is possible that the interaction of the Rik1-containing complex with Cdc20 underlies this silencing defect.

We reasoned that the loss of silencing in the mutant may be linked to the impaired DNA replication. Indeed, *cdc20-p7* cells grew much more slowly at 34 °C than at 23 °C and had an extended S phase (Fig. 1b). The replication state of a strain can be tested by assessing the efficiency of replication recovery after ultraviolet-radiation-induced damage. We found that the *cdc20-p7* mutant was highly sensitive to ultraviolet radiation at 34 °C but not at 23 °C (Fig. 2c). Furthermore, heterochromatin fragments that contain ARS elements could not be replicated efficiently in the mutant at 34 °C (Supplementary Fig. 11). Thus, the loss of heterochromatin silencing in the *cdc20-p7* mutant seems to be coupled to a defect in DNA replication.

To gain further insight into the role of Cdc20 in the heterochromatin pathway, we analysed an amino-terminal deletion of Cdc20, denoted *cdc20*^{AN-term}. The N terminus of Cdc20, which contains the catalytic domain, is not essential for cell survival²⁴. To test how this mutation affects heterochromatin silencing, the *cdc20*^{AN-term} mutant was crossed into the *otr::ade6*⁺ background and was analysed at 23 °C on rich medium that was not supplemented with extra adenine. In this system, WT cells formed red colonies, indicating transcriptional silencing, but the *cdc20*^{AN-term} mutant colonies were white (Fig. 2d), indicating that heterochromatin silencing is alleviated in the mutant. Consistent with this, H3K9 methylation in pericentromeric repeats was significantly reduced in the *cdc20*^{AN-term} mutant (Fig. 2e). Thus, DNA replication and heterochromatin function were decoupled in this mutant, further showing that Cdc20 is directly involved in heterochromatin silencing.

Here we have demonstrated that the Dos2-containing complex, which includes Dos2, Mms19, Rik1 and Cdc20, is crucial for DNA replication, siRNA production and heterochromatin assembly. Our findings establish the first physical and functional link between DNA replication, small RNA generation and H3K9 methylation, and they provide a novel mechanism to explain how these processes are coordinated (Fig. 4d).

Our results provide insight into how the epigenetic states of heterochromatin are accurately duplicated in each cell cycle (Fig. 4d). In budding yeast, heterochromatin assembly requires S-phase progression but not origin firing^{25,26}. Our findings suggest that, by contrast, DNA replication is required for heterochromatin assembly in fission yeast. In plants and mammals, DNA replication and DNA polymerase- ϵ have also been implicated in the silencing of heterochromatin^{27–29}. This finding suggests that a molecular mechanism linking DNA replication to

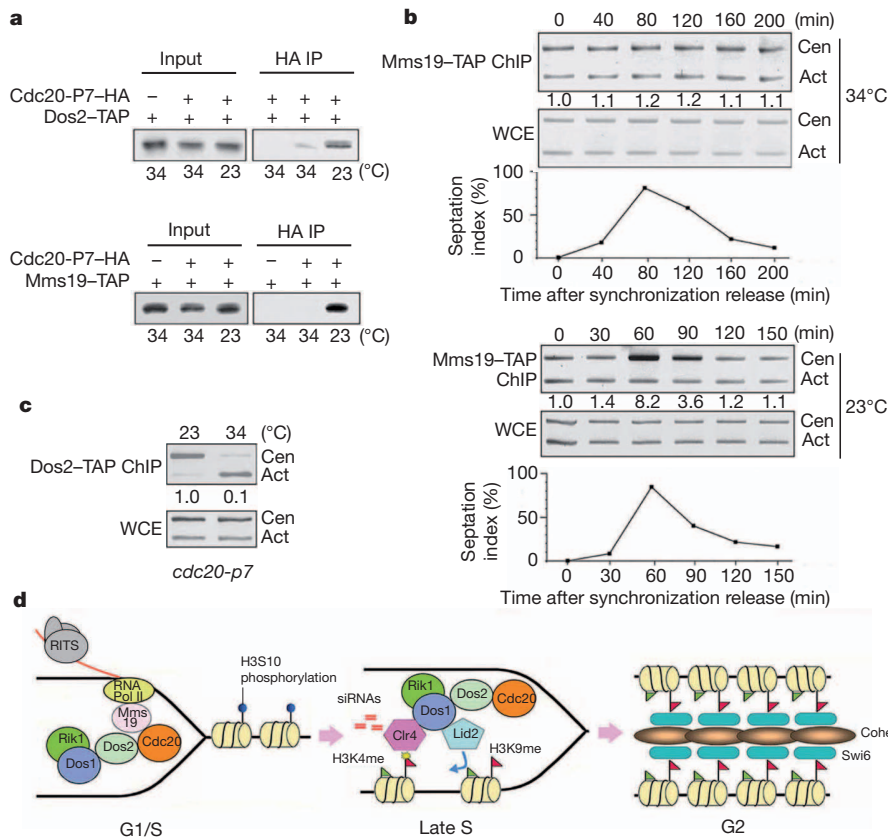


Figure 4 | Functional interactions between components of the Dos2-containing complex. **a**, The interaction of the *cdc20-p7* mutant with Dos2 or Mms19 was abolished at 34 °C but not at 23 °C. Co-immunoprecipitations (Co-IPs) for the two sets of strains (top and bottom) incubated at the indicated temperatures were performed using an HA-tagged antibody. Precipitates were analysed by western blotting with an antibody specific for the TAP tag. **b**, **c**, Mms19-TAP and Dos2-TAP accumulation in heterochromatin in S phase were lost in *cdc20-p7* cells at 34 °C. ChIP assays were performed on synchronized cells carrying either Mms19-TAP or Dos2-TAP by using an antibody specific for the TAP tag, followed by PCR, as in Fig. 3c. Cell cycle progression was monitored by a septation index. The relative fold enrichment of Mms19 or Dos2 is indicated below each lane. **d**, Model of how DNA

replication, small RNA generation and H3K9 methylation are coordinated. Left, during G1/S phase, while synthesizing the leading heterochromatin strands, Cdc20 regulates heterochromatin transcription by interacting with Mms19, and it also recruits Dos2 and Rik1. Another DNA polymerase subunit may be responsible for a similar process on the lagging strand. Heterochromatin transcripts are subsequently processed into siRNAs by the RNA-induced transcriptional silencing (RITS) complex. Centre, during late S phase, the complex containing Dos1, Dos2 and Rik1, together with the siRNAs, promotes H3K9 methylation by Ctr4. Right, during G2 phase, Swi6 binds to methylated H3K9 to reassemble chromatin into a repressed state and retain cohesin.

heterochromatin formation, similar to the one elucidated in this study, is probably conserved in multicellular eukaryotes.

METHODS SUMMARY

The *S. pombe* (fission yeast) strains used in this study are listed in Supplementary Table 1. Cell synchronization was performed by the hydroxyurea method. For mass spectrometry, TAP-tagged Dos2 was purified from a total of 9×10^{10} cells as described previously¹¹. Immunofluorescence images were taken using a DeltaVision Imaging System (Applied Precision). The program SoftWoRX 2.50 (Applied Precision) was used for processing the final projections. For the ultraviolet radiation survival assay, cells were collected from the culture at logarithmic phase and were plated at appropriate dilutions onto yeast extract with supplements (YES) medium. The plates were then irradiated with various doses of ultraviolet radiation. After incubation at 23 °C for 5 days, the colonies were counted. Detailed descriptions of the co-immunoprecipitation assays, ChIP assays, RT-PCR and northern blotting are provided in the Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 21 January; accepted 3 May 2011.

Published online 3 July 2011.

1. Kouzarides, T. Chromatin modifications and their function. *Cell* **128**, 693–705 (2007).

2. Allis, C. D., Jenuwein, T. & Reinberg, D. (eds) *Epigenetics* (Cold Spring Harbor Laboratory Press, 2006).
3. Probst, A. V., Dunleavy, E. & Almouzni, G. Epigenetic inheritance during the cell cycle. *Nature Rev. Mol. Cell Biol.* **10**, 192–206 (2009).
4. Wallace, J. & Orr-Weaver, T. Replication of heterochromatin: insights into mechanisms of epigenetic inheritance. *Chromosoma* **114**, 389–402 (2005).
5. Chen, E. S. *et al.* Cell cycle control of centromeric repeat transcription and heterochromatin assembly. *Nature* **451**, 734–737 (2008).
6. Kloc, A., Zaratiegui, M., Nora, E. & Martienssen, R. RNA interference guides histone modification during the S phase of chromosomal replication. *Curr. Biol.* **18**, 490–495 (2008).
7. Motamedi, M. R. *et al.* Two RNAi complexes, RITS and RDRC, physically interact and localize to noncoding centromeric RNAs. *Cell* **119**, 789–802 (2004).
8. Hong, E. J., Villen, J., Gerace, E. L., Gygi, S. P. & Moazed, D. A cullin E3 ubiquitin ligase complex associates with Rik1 and the Clr4 histone H3-K9 methyltransferase and is required for RNAi-mediated heterochromatin formation. *RNA Biol.* **2**, 106–111 (2005).
9. Horn, P. J., Bastie, J. N. & Peterson, C. L. A Rik1-associated, cullin-dependent E3 ubiquitin ligase is essential for heterochromatin formation. *Genes Dev.* **19**, 1705–1714 (2005).
10. Li, F. *et al.* Two novel proteins, Dos1 and Dos2, interact with Rik1 to regulate heterochromatic RNA interference and histone modification. *Curr. Biol.* **15**, 1448–1457 (2005).
11. Li, F. *et al.* Lid2 is required for coordinating H3K4 and H3K9 methylation of heterochromatin and euchromatin. *Cell* **135**, 272–283 (2008).
12. Thon, G. *et al.* The Clr7 and Clr8 directionality factors and the Pcu4 cullin mediate heterochromatin formation in the fission yeast *Schizosaccharomyces pombe*. *Genetics* **171**, 1583–1595 (2005).
13. White, S. A. & Allshire, R. C. RNAi-mediated chromatin silencing in fission yeast. *Curr. Top. Microbiol. Immunol.* **320**, 157–183 (2008).

14. Bannister, A. J. *et al.* Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* **410**, 120–124 (2001).
15. D'Urso, G. & Nurse, P. *Schizosaccharomyces pombe cdc20⁺* encodes DNA polymerase ϵ and is required for chromosomal replication but not for the S phase checkpoint. *Proc. Natl Acad. Sci. USA* **94**, 12491–12496 (1997).
16. Lauder, S. *et al.* Dual requirement for the yeast *MMS19* gene in DNA repair and RNA polymerase II transcription. *Mol. Cell. Biol.* **16**, 6783–6793 (1996).
17. Volpe, T. A. *et al.* Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* **297**, 1833–1837 (2002).
18. Hall, I. M., Noma, K. & Grewal, S. I. S. RNA interference machinery regulates chromosome dynamics during mitosis and meiosis in fission yeast. *Proc. Natl Acad. Sci. USA* **100**, 193–198 (2003).
19. Jia, S. T., Noma, K. & Grewal, S. I. S. RNAi-independent heterochromatin nucleation by the stress-activated ATF/CREB family proteins. *Science* **304**, 1971–1976 (2004).
20. Wu, X. Y., Li, H. & Chen, J. D. The human homologue of the yeast DNA repair and TFIIF regulator *MMS19* is an AF-1-specific coactivator of estrogen receptor. *J. Biol. Chem.* **276**, 23962–23968 (2001).
21. Ito, S. *et al.* MMXD, a TFIIF-independent XPD–MMS19 protein complex involved in chromosome segregation. *Mol. Cell* **39**, 632–640 (2010).
22. Natsume, T. *et al.* A DNA polymerase α accessory protein, Mcl1, is required for propagation of centromere structures in fission yeast. *PLoS ONE* **3**, e2221 (2008).
23. Nakayama, J., Allshire, R. C., Klar, A. J. S. & Grewal, S. I. S. Role for DNA polymerase α in epigenetic control of transcriptional silencing in fission yeast. *EMBO J.* **20**, 2857–2866 (2001).
24. Feng, W. Y. & D'Urso, G. *Schizosaccharomyces pombe* cells lacking the amino-terminal catalytic domains of DNA polymerase ϵ are viable but require the DNA damage checkpoint control. *Mol. Cell. Biol.* **21**, 4495–4504 (2001).
25. Kirchmaier, A. L. & Rine, J. DNA replication-independent silencing in *S. cerevisiae*. *Science* **291**, 646–650 (2001).
26. Li, Y. C., Cheng, T. H. & Gartenberg, M. R. Establishment of transcriptional silencing in the absence of DNA replication. *Science* **291**, 650–653 (2001).
27. Finnegan, E. J. & Dennis, E. S. Vernalization-induced trimethylation of histone H3 lysine 27 at FLC is not maintained in mitotically quiescent cells. *Curr. Biol.* **17**, 1978–1983 (2007).
28. Yin, H. *et al.* Epigenetic regulation, somatic homologous recombination, and abscisic acid signaling are influenced by DNA polymerase mutation in *Arabidopsis*. *Plant Cell* **21**, 386–402 (2009).
29. Fuss, J. & Linn, S. Human DNA polymerase ϵ colocalizes with proliferating cell nuclear antigen and DNA replication late, but not early, in S phase. *J. Biol. Chem.* **277**, 8658–8666 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. Allshire, P. Nurse, G. D'Urso and the Japan Yeast Genetic Resource Center for strains, E. Osborne and C. Hale for comments on the manuscript, Cold Spring Harbor Laboratory for mass spectrometry analysis, and members of the Cande and Li laboratories for their support and discussions. This work was supported by a grant from the National Institutes of Health (R01GM076396) to W.Z.C. and R.M.

Author Contributions F.L. and W.Z.C. designed the experiments and wrote the manuscript. F.L. performed the experiments. R.M. provided mass spectrometry expertise, equipment and conceptual support.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to F.L. (fl43@nyu.edu).

METHODS

Yeast strains, media and genetic procedures. The *S. pombe* (fission yeast) strains used in this study are listed in Supplementary Table 1. Yeast extract with supplements (YES) was used as a complete culture medium, Edinburgh minimal medium (EMM) as a minimal medium, and SPAS medium for conjugation and sporulation. Cell synchronization was performed by the hydroxyurea method. Briefly, cells were treated with 12 mM hydroxyurea for 4 h and then released into 100 µg ml⁻¹ thiabendazole for 1 h to block entry to mitosis. Standard genetic protocols for fission yeast were used³⁰.

Mass spectrometry. TAP-tagged Dos2 was purified from a total of 9×10^{10} cells as described previously¹¹. Briefly, cell lysates in 1× lysis buffer (50 mM bis-Tris propane, pH 7.0, 0.1 M KCl, 5 mM EDTA, 5 mM EGTA and 10% glycerol) were incubated with IgG sepharose (Amersham Pharmacia Biotech) for 2 h. After washing with lysis buffer, the IgG sepharose was incubated overnight with TEV protease (Invitrogen). The supernatant was removed from the IgG sepharose and added to S-protein agarose slurry (Novagen) for 3 h. The S-protein agarose was then washed with lysis buffer. The eluate from this agarose was analysed by silver staining and subjected to mass spectrometry (at Cold Spring Harbor Laboratory).

Immunoprecipitation assays. Cells were lysed in HB buffer by using the glass bead method³⁰. Lysates were pre-cleared with protein A agarose beads, which was followed by a 2-h incubation with anti-HA or anti-GFP antibody (Sigma) or IgG Sepharose 6 Fast Flow beads (Amersham Biosciences) at 4 °C. After washing, the eluted proteins and input extracts were analysed by western blotting using anti-S-tag (MA1-981, ABR Affinity BioReagents) or anti-RNA Pol II (ab5408, Abcam) antibody.

ChIP analysis. CHIP assays were carried out as described previously³¹. Cells taken from culture at logarithmic phase were crosslinked with 1% formaldehyde. Immunoprecipitation was performed with S-protein agarose (Novagen) or antibodies specific for the following: dimethylated H3K9 (07-441, Upstate) or Swi6 (ab14898, Abcam). The precipitated DNA was analysed by using competitive PCR with oligonucleotides specific for the centromeric *dh* region or a control gene, *act1*⁺. The PCR products were separated on a 1.7% agarose gel and post-stained with ethidium bromide. The primers used are listed in Supplementary Table 2.

RT-PCR. Total RNA was isolated from cells taken from culture growing at logarithmic phase by using an RNeasy Mini Kit (Qiagen). After treatment with DNase I (Promega), 50 ng purified RNA was analysed by RT-PCR in a 25 µl

reaction volume using a one-step RT-PCR kit (Qiagen). Equal loading of RNA samples was assessed by amplification of a control gene, *act1*⁺. For strand-specific RT-PCR, RNA samples were incubated with primers that were complementary to the forward or the reverse centromeric transcripts for synthesis of the first cDNA strand. After heat inactivation of the reverse transcriptase at 95 °C for 15 min, a second primer was added for the subsequent cycles of PCR amplification. The primers used for RT-PCR are listed in Supplementary Table 2.

Small RNA northern blotting analyses. Small RNA northern blotting analysis was performed as described previously¹⁰. Briefly, siRNAs were extracted from cells taken when the culture was in logarithmic phase, in YES medium, using a mirVana miRNA Isolation Kit (Ambion). Total small RNA (25 µg) was resolved on a 15% denaturing acrylamide gel and was blotted to a charged nylon membrane (Hybond-N+, Amersham). RNA blots were crosslinked and hybridized with DNA probes specific for the centromeric *otr* region or snoR69 as a loading control. The DNA oligonucleotides used as probes are listed in Supplementary Table 2.

Microscopy. Fluorescent images were taken using a DeltaVision Imaging System (Applied Precision). The program SoftWoRX 2.50 (Applied Precision) was used for processing the final projections. The standard procedure was followed by staining with 4',6-diamidino-2-phenylindole (DAPI).

Ultraviolet radiation survival assay. Cells were collected when the culture was in logarithmic phase and were plated at appropriate dilutions onto YES medium. The plates were irradiated with various doses of ultraviolet radiation. After incubation at 23 °C for 5 days, the colonies were counted.

DNA replication assay. A pBluescript plasmid carrying a 3.0-kb fragment from heterochromatic *cen3* repeats (3.0-K)³², together with a *ura4*⁺ selective marker, was transformed into the *cdc20-p7* mutant. This centromeric fragment contains efficient ARS elements³². After incubation at 23 °C or 34 °C for 4 days, total DNA was isolated following standard procedures³². Purified DNA (50 ng) was analysed by PCR (Bio-Rad) with primers specific for *ura4*⁺. Equal loading was assessed by amplification of a control gene, *act1*⁺.

30. Moreno, S., Klar, A. & Nurse, P. Molecular genetic analysis of fission yeast *Schizosaccharomyces pombe*. *Methods Enzymol.* **194**, 795–823 (1991).
31. Pidoux, A., Mellone, B. & Allshire, R. Analysis of chromatin in fission yeast. *Methods* **33**, 252–259 (2004).
32. Smith, J. G. *et al.* Replication of centromere II of *Schizosaccharomyces pombe*. *Mol. Cell. Biol.* **15**, 5165–5172 (1995).

Architecture of the Mediator head module

Tsuyoshi Imasaki¹, Guillermo Calero^{2†}, Gang Cai^{3†}, Kuang-Lei Tsai³, Kentaro Yamada¹, Francesco Cardelli¹, Hediye Erdjument-Bromage⁴, Paul Tempst⁴, Imre Berger⁵, Guy Lorch Kornberg², Francisco J. Asturias³, Roger D. Kornberg² & Yuichiro Takagi¹

Mediator is a key regulator of eukaryotic transcription¹, connecting activators and repressors bound to regulatory DNA elements with RNA polymerase II^{1–4} (Pol II). In the yeast *Saccharomyces cerevisiae*, Mediator comprises 25 subunits with a total mass of more than one megadalton (refs 5, 6) and is organized into three modules, called head, middle/arm and tail^{7–9}. Our understanding of Mediator assembly and its role in regulating transcription has been impeded so far by limited structural information. Here we report the crystal structure of the essential Mediator head module (seven subunits, with a mass of 223 kilodaltons) at a resolution of 4.3 ångströms. Our structure reveals three distinct domains, with the integrity of the complex centred on a bundle of ten helices from five different head subunits. An intricate pattern of interactions within this helical bundle ensures the stable assembly of the head subunits and provides the binding sites for general transcription factors and Pol II. Our structural and functional data suggest that the head module juxtaposes transcription factor IIH and the carboxy-terminal domain of the largest subunit of Pol II, thereby facilitating phosphorylation of the carboxy-terminal domain of Pol II. Our results reveal architectural principles underlying the role of Mediator in the regulation of gene expression.

In the yeast *S. cerevisiae*, the Mediator head module is composed of seven subunits¹⁰: Med17 (also known as Srb4), Med11, Med22 (Srb6), Med6, Med8, Med18 (Srb5) and Med20 (Srb2). Four subunits are encoded by *SRB* genes, first identified through a genetic screen for mutations suppressing the Pol II carboxy-terminal domain (CTD) truncation^{11,12}. The head module is essential for Mediator function because mutations in the head abolish messenger RNA synthesis *in vivo*^{13,14} and *in vitro*¹⁵, and eliminate Mediator interaction with promoters *in vivo*¹⁰. The head module is organized into three domains that can undergo significant conformational changes, and it interacts with the TATA-binding protein subunit of general transcription factor TFIID and the Rpb4 and Rpb7 subunits of Pol II (ref. 16). The head has also been shown to interact with TFIIF through the Med11 subunit¹⁷. Determining the architecture of the Mediator head module is therefore vital to understanding the mechanism by which Mediator controls gene expression.

We engineered the head module to obtain crystals of sufficient quality for structure determination (Supplementary Information, section 1). In our engineered Mediator head, Med18 loop regions and the amino-terminal 108 residues of Med17 were deleted, without apparent effect on the integrity of the complex (Supplementary Fig. 1). The modified head module was labelled with selenomethionine (SeMet) and purified as described previously¹⁶. By overcoming two major technical obstacles (Supplementary Information, section 2), we produced SeMet crystals that diffract to 4.3 Å (Supplementary Table 1). The electron density map was calculated to a resolution of 4.3 Å (Supplementary Fig. 2) by SeMet single anomalous dispersion (SAD) after initial phases had been obtained using Ta₆Br₁₄ and K₃Ir(NO₃)₆ derivatives (Supplementary Information, section 3).

We began identification of the individual polypeptide constituents of the Mediator head module by docking the Med18–Med20–Med8 C-terminal helix (CTH) complex structure¹⁸ (Protein Data Bank ID, 2HZS) into the electron density map and then performing rigid-body refinement. The polypeptide chains of the other subunits were identified on the basis of the SeMet positions and their juxtaposition with large amino-acid side chains within ordered regions of secondary structure (Methods). This approach permitted the unambiguous assignment of all discernible elements of secondary structure in the density map to individual head module subunits (Fig. 1 and Supplementary Figs 3–5).

Our crystal structure is consistent with the molecular envelope of the head module derived at a resolution of 30–35 Å by single-particle electron microscopy analysis (Supplementary Figs 6 and 7). The head can be described in terms of three major domains, a ‘fixed jaw’, a ‘movable jaw’ and a ‘neck’ (Fig. 1 and Supplementary Figs 4 and 5), with a ‘central joint’ connecting these domains. Our X-ray structure of the head module reveals the overall architecture of the module and the domain boundaries. The domains are connected through flexible loops and linkers at the central joint.

Our previous work on expression and purification of the head module suggested that Med17 has a central role in head assembly¹⁰. The work we report here extends those results through a comprehensive biochemical analysis in combination with electron microscopy, to determine the Med17 domain structure and elucidate its interactions with other head components (Supplementary Information, section 4). The results support our model of the architecture of the head module.

Assembly of the head module starts with formation of the ‘mini-head’ (Med17–Med11–Med22). Subsequently, Med8 and Med6 are added, followed by Med20–Med18 (ref. 10). Our structure shows that a four-helix bundle, built by α -helices from Med11 (BH1 and BH2) and Med22 (BH1 and BH2) interact with BH2 of Med17 to form the larger helical bundle (Figs 2 and 3 and Supplementary Fig. 4). This is consistent with the observation that omission of either Med11 or Med22 leads to disassembly of the head¹⁰. Med6 interacts with the mini-head through its BH1, and Med8 serves to stabilize the central α -helical bundle by surrounding the central helices. Finally, the Med18–Med20 heterodimer binds to the core-head, which is composed of five subunits (Med6, Med8, Med11, Med22 and Med17), primarily through the CTH of Med8 (Fig. 2).

The fixed jaw domain comprises the CTHs of Med11 and Med22 and the CTD of Med17. The Med11 and Med22 CTHs interact with the helical regions of the Med17 CTD. Med17 (residues 610–660) forms a β -sheet structure that lines the inner surface of the fixed jaw and faces the movable jaw (Fig. 3a). The Med17 CTD interacts with the loop region of Med18. The functional importance of the Med17 CTD correlates with the biochemical activity of the Head module *in vitro*, as well as phenotypic analysis *in vivo*, as loss of the Med17 CTD abolishes the transcription activity of the head module (Supplementary Fig. 12),

¹Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, 635 Barnhill Drive, Indianapolis, Indiana 46202, USA. ²Department of Structural Biology, Stanford University School of Medicine, Stanford, California 94350, USA. ³Department of Cell Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA. ⁴Molecular Biology Program, Memorial Sloan-Kettering Cancer Center, New York, New York 10021, USA. ⁵European Molecular Biology Laboratory, Grenoble Outstation, 6 rue Jules Horowitz, 38042 Grenoble Cedex 9, France. [†]Present addresses: Department of Structural Biology, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15260, USA (G. Calero); School of Life Sciences, University of Science and Technology of China, Hefei, Anhui 230027, China (G. Cai).

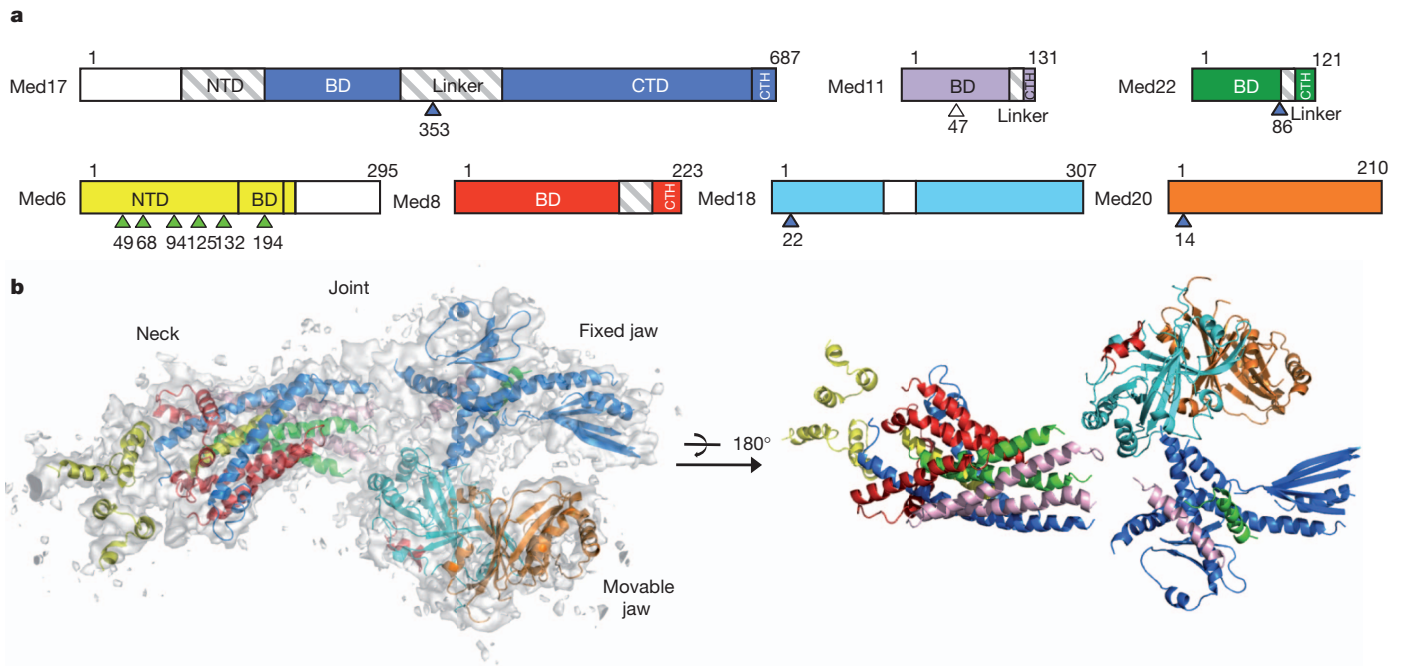


Figure 1 | Overall structure of the Mediator head module. **a**, Head module subunit domains. Med17 is shown in blue, Med11 in purple, Med22 in dark green, Med6 in yellow, Med8 in red, Med18 in cyan and Med20 in orange. The regions not modelled are hatched in grey and the regions not present in the crystal are shown in white. Positions of *med6^{ts}* mutations are marked by green

arrows, *srb* suppressor mutations by blue arrows and Med11 residue 47 (Thr) by a white arrow. BD, bundle domain; CTH, C-terminal domain; NTD, N-terminal domain. **b**, A ribbon model of the Mediator head module superimposed on the experimental electron density map contoured at 1.5 σ .

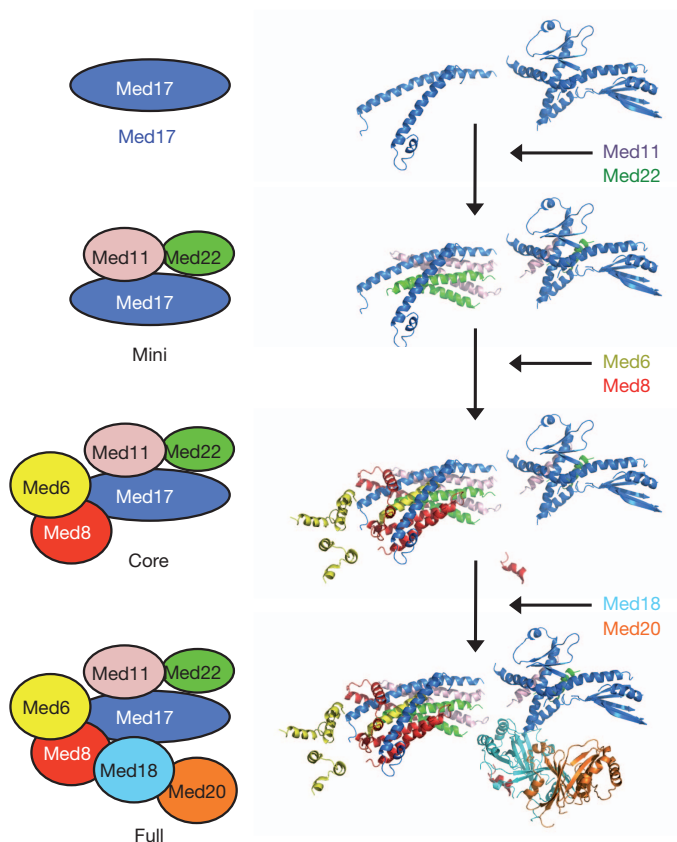


Figure 2 | Mechanism of Mediator head module complex assembly. Models of the mini-head (Med17, Med11 and Med22) and core-head (mini-head with Med6 and Med8) modules as derived from our crystal structure of the full head module (core-head with Med18 and Med20). Diagrams of head module components (left) and corresponding structures (right) are shown.

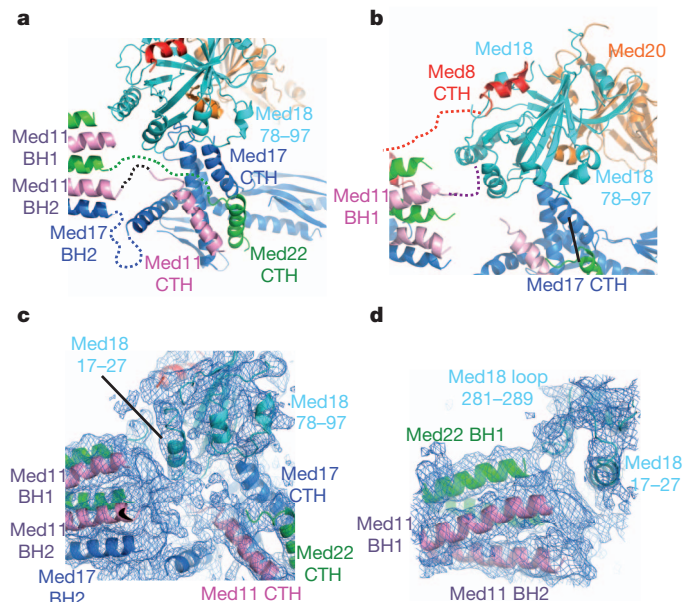


Figure 3 | Structures of fixed and movable jaw domains. **a**, Fixed jaw domain interactions. The linker regions of Med17 (residues 320–420), Med11 (94–110) and Med22 (87–100) are drawn as dotted lines. **b**, Movable jaw domain interactions. Linker regions of Med8 and Med11 are drawn as dotted lines. **c**, Electron density map at the central joint region, showing density corresponding to the linker regions of Med11, Med22 and Med17. **d**, Electron density map at the junction of the Med11, Med22 and Med18 subunits. The models of Med22 BH1, Med11 BH1 and BH2, and the Med18 loop region (residues 17–27 and 281–289) are superimposed.

and all Med17 CTD deletion mutants as well as internal deletion mutants result in lethality (Supplementary Fig. 13).

The movable jaw, so called because previous electron microscopy studies¹⁶ demonstrated multiple orientations of this domain with respect to the rest of the head module, is formed by the Med18–Med20–Med8 CTH complex. As for the interaction with the Med8 CTH¹⁸, our complete head module structure has revealed additional interactions with the fixed jaw and neck domains. First, the Med18 loop region formed by residues 78–97 interacts with the Med17 CTH of the fixed jaw domain (Fig. 3a, b). Second, the electron density corresponding to the N terminus of the Med11 subunit (residues 1–20) indicates an interaction with Med18 (residues 17–27 and 281–289; Fig. 3c, d). The assignment of Med11 residues 1–20 was complicated by the substitution of Ser 17 for Met 17 (Methods), and an unambiguous sequence marker is therefore lacking. However, our biochemical data (Supplementary Information, section 5) support our architectural model, in which a stable association between Med18–Med20 and the head module requires binding to Med8 and at least one additional interaction (with Med11 or Med17). The interactions with the CTH of Med17 and the NTD of Med11 are likely to be critical for the functional positioning and flexibility of the movable jaw¹⁶ (Fig. 3b–d and Supplementary Fig. 6).

The neck domain has an unusual structure: a total of ten helices from five different subunits associate through the formation of a large helical bundle. The NTD of Med6 is located adjacent to the large helical bundle and consists of four α -helices (Figs 1b and 4a). The helical bundle of the neck domain can be divided into two parts, a

short bundle composed of four short α -helices and a long bundle composed of six long α -helices. Three helices of the Med8 subunit (BH3, BH4 and BH5) seem to stabilize the assembly of both short and long bundles, and, thus, the entire neck domain structure. TATA-binding protein was reported to bind to the N-terminal 138 residues of Med8 (ref. 18), which corresponds to helices BH1 to BH5, all of which are located on the surface of the neck domain.

The organization of the helical bundle in the neck domain may produce a relatively rigid structure that could mechanically convey regulatory signals. Several observations suggest that Med6 may function as an interface between the Mediator head and middle modules, and transduce a mechanical signal from the tail or middle to the head and onto Pol II (Supplementary Information, section 6).

Mediator stimulates the phosphorylation of Ser 5 in the Pol II CTD by TFIIF (ref. 19), which promotes dissociation of Mediator from Pol II (refs 20, 21), an important step in the transition from initiation to elongation of transcription²². Our structural and biochemical data, along with relevant previous observations^{12,17,23,24}, suggest an interaction of the Pol II CTD, the Mediator head module and TFIIF. First, mutation of Thr 47 to Ala in Med11 affects the interaction of TFIIF with the head module *in vivo*, resulting in a reduction of Pol II CTD Ser 5 phosphorylation¹⁷. Thr 47 of Med11 is located near the centre of the two symmetrical, long helical bundles of the neck, which thus could constitute the docking surface for TFIIF (Fig. 4b and Supplementary Fig. 16c). Second, three of four suppressor mutations of Pol II CTD truncation—Med17 (Gly 353 to Cys), Med22 (Asn 86 to Lys) and Med18 (Thr 22 to Ile)—map to the central joint region^{12,24} (Supplementary Fig. 16a), suggesting that there is a functional interaction between the CTD and this portion of the head, consistent with previous observations²³. Third, the head module within the Mediator/Pol II structure (Fig. 4b and Supplementary Fig. 17) is located near the base of the CTD. Finally, our biochemical data show that the head module stimulates phosphorylation of the Pol II CTD by TFIIF (Supplementary Information, section 7, and Supplementary Fig. 18). Therefore, we suggest that the head module may function as a scaffold that juxtaposes TFIIF and the Pol II CTD, thereby facilitating CTD phosphorylation (Fig. 4b). Our Mediator head structure reveals intricate interaction networks, notably the striking multi-helical bundle in the neck domain, engaging five Mediator subunits in a single structure unit. Such interactions could not have been determined from structures of individual subunits alone, nor from analysing pairwise small domain–domain interactions, but only by study of the multi-protein complex in its entirety.

METHODS SUMMARY

Structure determination. Modified head module was expressed with the MultiBac system²⁵ in insect cells and purified by nickel affinity chromatography. Crystals were obtained by the hanging-drop vapour diffusion method. The structure was determined by SeMet SAD after a sufficient number (98) of SeMet sites had been identified from a combination of initial phases obtained using Ta₆Br₁₄ and iridium derivatives and partial-model SAD phases.

Biochemical and electron microscopy analysis. The Mediator head and its mutants were expressed in insect cells and purified by nickel affinity chromatography. The electron microscope images of the head module and the mutants were collected and class averages were calculated.

In vitro assays and yeast genetics. The *in vitro* transcription assay to assess activity of the recombinant head module and its mutant form using *srb4^{ts}* mutant crude extract, the assay for phosphorylation of the CTD of Pol II by TFIIF, and the yeast phenotypic analysis were all done as described previously¹⁵.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 24 January; accepted 28 April 2011.

Published online 3 July 2011.

1. Kornberg, R. D. Mediator and the mechanism of transcriptional activation. *Trends Biochem. Sci.* **30**, 235–239 (2005).

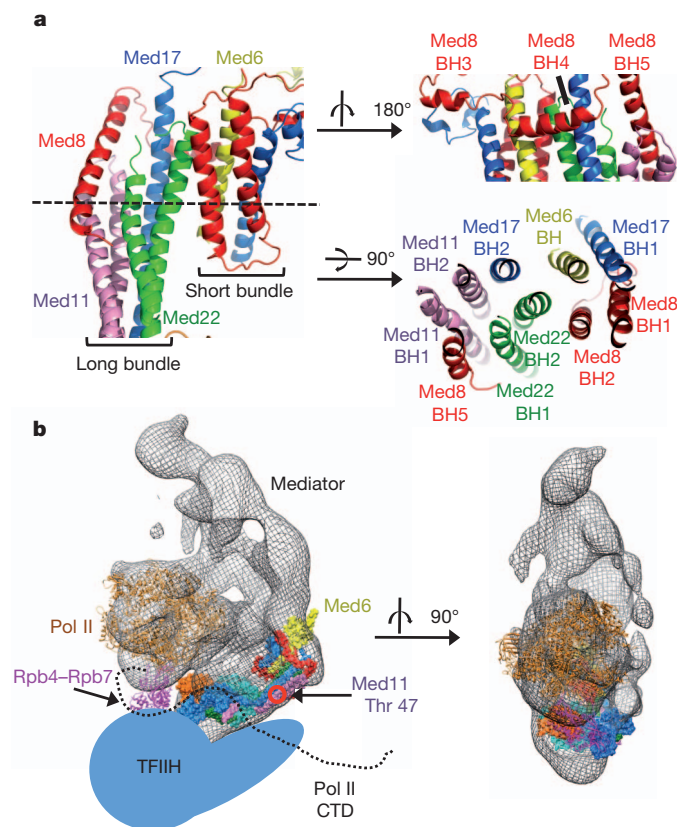


Figure 4 | Structure of the neck domain, and model of the Pol II–Mediator–TFIIF complex. **a**, The neck domain is depicted in front (left), back (top right) and top (bottom right) views. **b**, Model of the Pol II–Mediator–TFIIF complex. Pol II and the head structures were docked into the electron microscopy map of Mediator–Pol II, shown as a mesh⁸. The head module is coloured as in Fig. 1b. Core Pol II is in brown, Rpb4–Rpb7 is in purple, the Pol II CTD is drawn as a black dotted line and TFIIF is shown schematically (light blue). The location of Med11 residue 47 (Thr) is indicated.

2. Conaway, R. C., Sato, S., Tomomori-Sato, C., Yao, T. & Conaway, J. W. The mammalian Mediator complex and its role in transcriptional regulation. *Trends Biochem. Sci.* **30**, 250–255 (2005).
3. Boube, M., Joulia, L., Cribbs, D. L. & Bourbon, H. M. Evidence for a mediator of RNA polymerase II transcriptional regulation conserved from yeast to man. *Cell* **110**, 143–151 (2002).
4. Malik, S. & Roeder, R. G. The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. *Nature Rev. Genet.* **11**, 761–772 (2010).
5. Bjorklund, S. & Gustafsson, C. M. The yeast Mediator complex and its regulation. *Trends Biochem. Sci.* **30**, 240–244 (2005).
6. Guglielmi, B. *et al.* A high resolution protein interaction map of the yeast Mediator complex. *Nucleic Acids Res.* **32**, 5379–5391 (2004).
7. Asturias, F. J., Jiang, Y. W., Myers, L. C., Gustafsson, C. M. & Kornberg, R. D. Conserved structures of mediator and RNA polymerase II holoenzyme. *Science* **283**, 985–987 (1999).
8. Davis, J. A., Takagi, Y., Kornberg, R. D. & Asturias, F. A. Structure of the yeast RNA polymerase II holoenzyme: Mediator conformation and polymerase interaction. *Mol. Cell* **10**, 409–415 (2002).
9. Cai, G., Imasaki, T., Takagi, Y. & Asturias, F. J. Mediator structural conservation and implications for the regulation mechanism. *Structure* **17**, 559–567 (2009).
10. Takagi, Y. *et al.* Head module control of mediator interactions. *Mol. Cell* **23**, 355–364 (2006).
11. Nonet, M. L. & Young, R. A. Intragenic and extragenic suppressors of mutations in the heptapeptide repeat domain of *Saccharomyces cerevisiae* RNA polymerase II. *Genetics* **123**, 715–724 (1989).
12. Thompson, C. M., Koleske, A. J., Chao, D. M. & Young, R. A. A multisubunit complex associated with the RNA polymerase II CTD and TATA-binding protein in yeast. *Cell* **73**, 1361–1375 (1993).
13. Thompson, C. M. & Young, R. A. General requirement for RNA polymerase II holoenzymes *in vivo*. *Proc. Natl Acad. Sci. USA* **92**, 4587–4590 (1995).
14. Holstege, F. C. *et al.* Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* **95**, 717–728 (1998).
15. Takagi, Y. & Kornberg, R. D. Mediator as a general transcription factor. *J. Biol. Chem.* **281**, 80–89 (2006).
16. Cai, G. *et al.* Mediator head module structure and functional interactions. *Nature Struct. Mol. Biol.* **17**, 273–279 (2010).
17. Esnault, C. *et al.* Mediator-dependent recruitment of TFIIF modules in preinitiation complex. *Mol. Cell* **31**, 337–346 (2008).
18. Larivière, L. *et al.* Structure and TBP binding of the Mediator head subcomplex Med8–Med18–Med20. *Nature Struct. Mol. Biol.* **13**, 895–901 (2006).
19. Kim, Y. J., Bjorklund, S., Li, Y., Sayre, M. H. & Kornberg, R. D. A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell* **77**, 599–608 (1994).
20. Svejstrup, J. Q. *et al.* Evidence for a mediator cycle at the initiation of transcription. *Proc. Natl Acad. Sci. USA* **94**, 6075–6078 (1997).
21. Max, T., Sogaard, M. & Svejstrup, J. Q. Hyperphosphorylation of the C-terminal repeat domain of RNA polymerase II facilitates dissociation of its complex with mediator. *J. Biol. Chem.* **282**, 14113–14120 (2007).
22. Payne, J. M., Laybourn, P. J. & Dahmus, M. E. The transition of RNA polymerase II from initiation to elongation is associated with phosphorylation of the carboxyl-terminal domain of subunit IIa. *J. Biol. Chem.* **264**, 19621–19629 (1989).
23. Kang, J. S. *et al.* The structural and functional organization of the yeast mediator complex. *J. Biol. Chem.* **276**, 42003–42010 (2001).
24. Koleske, A. J., Buratowski, S., Nonet, M. & Young, R. A. A novel transcription factor reveals a functional link between the RNA polymerase II CTD and TFIID. *Cell* **69**, 883–894 (1992).
25. Fitzgerald, D. J. *et al.* Protein complex expression by using multigene baculoviral vectors. *Nature Methods* **3**, 1021–1032 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank L. Messerle for providing the Ta₆Br₁₄ metal cluster, T. Hurley for reading the manuscript, M. Georgiadis for discussions, C. Kaplan for his advice on yeast genetics, T. Earnest for giving us beam time and L. Fabrizio for assisting with the N-terminal sequence analysis. We thank the CCP4 summer school, funded by the NCI (Y1-CO-1020), and NIGMS (Y1-GM-1104) for their assistance with the twinning data analysis. Y.T. thanks the instructors on 'The X-Ray Methods Course' at Cold Spring Harbor Laboratory. This work was supported by US National Science Foundation grant MCB 0843026 (Y.T.); the American Heart Association 0735395N (Y.T.); a Human Frontier Science Program long-term fellowship (T.I.); NIH grants R01 GM67167 (F.J.A.) and GM36659 (R.D.K.); NCI Cancer Center Support Grant P30 CA08748 (to the MSKCC Microchemistry and Proteomics Core Laboratory); and the European Commission Framework Program 7 projects INSTRUCT and P-CUBE (I.B.). X-ray data were collected at the GM/CA-CAT at the Advanced Photon Source, Argonne National Laboratory. GM/CA-CAT is funded by the NIH (Y1-CO-1020 and Y1-GM-1104) and the Advanced Photon Source is supported by the DOE (DE-AC02-06CH11357). Portions of this research were carried out at the Stanford Synchrotron Radiation Lightsource, supported by the DOE and the NIH, and at the Advanced Light Source, supported by the DOE (DE-AC02-05CH11231).

Author Contributions T.I., I.B. and Y.T. implemented the MultiBac system. T.I. was mainly responsible for protein complex preparation, crystallization, data collection, data analysis and model building in collaboration with Y.T. T.I., H.E.-B. and P.T. carried out mass spectroscopy analysis. G. Calero, G.L.K. and Y.T. carried out the initial crystallization and data collection, supervised by R.D.K.. Y.T., T.I. and F.C. designed and carried out expression of the mutant head modules and their biochemical characterization. Y.T. and K.Y. designed and carried out the yeast genetic experiment. Y.T. carried out the *in vitro* transcription assay and the CTD kinase assay; G. Cai, K.-L.T. and F.J.A. carried out the electron microscopy study on the head module and its mutants. T.I., F.J.A. and Y.T. discussed and interpreted all results. Y.T. supervised the X-ray, biochemical and yeast genetic work, and wrote the manuscript in collaboration with T.I., I.B., F.J.A. and R.D.K.

Author Information Coordinates and structure factors have been deposited in the Protein Data Bank under accession code 3RJ1. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to Y.T. (ytakagi@iupui.edu).

METHODS

Construction of vectors. All the vectors used in this study are summarized in Supplementary Tables 3 and 4. For expression of the modified head module for crystallization, DNA sequences corresponding to residues 1–108 of Med17 were removed from vector pFL-10xHis-Med17 (ref. 16; pYT49) by the SLIC method²⁶, yielding pFL-10xHis-Med17 (109–687) (pYT171). DNA sequences corresponding to residues 109–140 and 71–156 of Med18, respectively, were removed from vector pSPL-Med18-Med20 (ref. 16; pYT75), yielding pSPL-Med20-Med18 (Δ 109–140) (pYT115) and pSPL-Med20-Med18 (Δ 71–156) (pYT114). To eliminate an alternative translation start site, the Met (residue 17) of Med11 was mutated to Ser (pYT311). Finally, the transfer vector for the modified head module was generated by fusing three vectors, pYT171, pYT114 and pUCDM-Med6-Med22-Med11-Med8 (pYT120) by Cre/LoxP recombination as previously described²⁵.

DNA sequences corresponding to residues 1–16 of Med11 were removed from the vector pYT111 by SLIC, yielding the vector pUCDM-Med22-Med11 (Δ 1–16) (pYT147). Fusion of pYT171, pYT147 and pYT120 with either pYT114 or, alternatively, pYT115 generated the expression vectors for a series of double Med18–Med11 partial head module deletion mutants.

The constructs for Med17 mutagenesis were generated as follows. BamHI and HindIII fragments corresponding to the C-terminal deletion mutants of Med17 were generated by first introducing a stop codon and a HindIII site (TAAAGCTT) into pBacPAK9-10His-SRB4 (MED17) vector¹⁰ adjacent to the sequences corresponding to residues 108, 200, 300, 400, 500 and 600 of Med17, by using the QuickChange mutagenesis kit (Stratagene), followed by BamHI and HindIII digestion and gel purification. The respective purified fragments were cloned into the BamHI and HindIII sites of pFL vector²⁵, yielding vectors pYT165 to pYT170 (Supplementary Table 3). The N-terminal deletion, as well as the internal deletion mutant constructs of Med17, were generated by removing DNA sequences corresponding to residues 1–108, 1–201, 1–302, 1–400, 101–200, 201–300 and 301–400 from pFL-10xHis-Med17 by the SLIC method, yielding the respective vectors pYT183 to pYT186 and pYT289 to pYT291 (Supplementary Table 3). These vectors were fused with pUCDM-Med6-Med22-Med18-Med20-Med11-Med8 (pYT151), yielding vectors encoding for head modules comprising Med17 mutant forms. The vector pYT151 was created by two rounds of sequential cloning of PmeI and the AvrII fragments containing Med18–Med20 and Med22–Med11 into SpeI and NruI sites of pUCDM-Med6-Med8 (pYT110).

Introduction of the deletion mutations into the yeast shuttle vector pCT127, carrying the wild-type MED17 gene, was also carried out by SLIC. The yeast shuttle vectors used in this study are listed in Supplementary Table 4.

Expression and purification of the head module and mutants. Expression and purification of the recombinant head module, the mutant forms and the subcomplex was carried out in insect cells using the MultiBac system²⁵. Production of high-titer viruses in Sf9 cells, and expression and purification of recombinant head modules of Mediator and its mutant forms was carried out as described previously¹⁶.

Preparation of SeMet labelled Head module will be described elsewhere (T.I. *et al.*, manuscript in preparation). Briefly, the insect cells were cultured in Met-free medium (Expression Systems) overnight before baculovirus infection. L-selenomethionine (20 mg l^{−1}; Sigma-Aldrich) was added at sequential 24-h intervals. Cells were collected 96 h after infection. SeMet-labelled complex was purified as described above.

Limited proteolysis and identification of the peptide fragments. A total of 135 μ g of the recombinant head module was incubated at 37 °C with chymotrypsin (Sigma-Aldrich) at a final concentration of 0.01 mg ml^{−1} in a volume of 150 μ l in buffer A (20 mM Tris-HCl (pH 8.0), 100 mM NaCl and 1 mM DTT). Aliquots (20 μ l) were taken at 0, 5, 10, 30 and 60 min, and 15 μ l of PMSF stock solution (100 mg ml^{−1}) was added to stop the reaction by inhibiting the protease. Aliquots were applied to 12.5% SDS–PAGE and transferred onto a Sequi-Blot PVDF membrane (Bio-Rad). Protein bands were stained by Coomassie blue (R-250). Protein bands resulting from proteolysis during the time course were identified, excised and subjected to Edman degradation using a Procise 494 instrument from Applied Biosystems as previously described²⁷. Stepwise-liberated PTH-amino acids were identified using an ‘on-line’ HPLC system (Applied Biosystems) equipped with a PTH C18 (2.1 \times 220 mm; 5- μ m particle size) column (Applied Biosystems).

Crystallization and data collection. Crystals were obtained at 293 K by hanging-drop vapour diffusion against a reservoir solution of 0.1 M Tris-HCl (pH 7.6) containing 10–12.5% (w/v) PEG-6K and 0.4 M (NH₄)₂SO₄. Crystals were transferred into the reservoir solution containing 25% triethylene glycol (TEG). The crystals were flash-frozen for data collection at 100 K. SDS–PAGE analysis of the dissolved crystals confirmed the presence of all seven subunits. However, *in situ* proteolysis resulted in about 10% of the Med17 subunits being shortened at the N terminus by 76 residues and almost 100% of the Med6 subunits being shortened at the C terminus by 80 residues (Supplementary Fig. 1). Diffraction data were collected at beamline 23ID

at the Advanced Photon Source (APS) at Argonne National Laboratory. All diffraction data were processed with HKL2000²⁸. Twinning rates of the data sets were analysed using program PHENIX XTRIAGE²⁹.

Structure determination of the Mediator head module. Initial phases were determined by two approaches: Ta₆Br₁₄ single isomorphous replacement with anomalous scattering (SIRAS) and iridium single anomalous dispersion (SAD). Ta₆Br₁₄ derivative crystals were prepared by soaking the native head module crystals in reservoir solution containing 1 mM Ta₆Br₁₄. The initial phase was determined by SIRAS at a resolution of 7.5 Å. Density modification using the program PARROT extended the phase resolution to 4.3 Å using the SeMet data set. Iridium derivatives of the crystal of Mediator head module were prepared by soaking the crystals in crystallization reservoir solution containing 10 mM K₃Ir(NO₃)₆. The initial iridium phase was obtained by SAD using the programs SHELXD and PHASER^{30,31}. The phase was extended followed by density modification by program PARROT³² with the SeMet data set. However, the maps obtained at this stage were not yet interpretable.

To improve the maps, we used them together and applied the following methods: (i) location of SeMet sites in the crystal; (ii) non-crystallographic symmetry (NCS), averaging between three molecules in the NCS using the program DM³³; (iii) partial model building into the clearly discernible rod-like electron density from α -helices, followed by rigid-body refinement using the programs COOT and REFMAC5³⁴; and (iv) re-calculating phases by SeMet SAD phasing with PHASER, using the partial model and SeMet positions. Iterative rounds combining these procedures were performed until the model covered all interpretable secondary structure elements. Eventually, we could identify 98 SeMet sites. To minimize model bias, phases were re-calculated by SeMet SAD with PHASER, using only the positions of these 98 selenium sites, and these improved SAD phases guided the final model-building steps.

Model building and refinement. Assignment of polypeptide identities was carried out as follows. The published structure¹⁸ of Med18–Med20–Med8 (CTH) (PDB ID, 2HZS) was manually docked into the electron density map, followed by rigid-body refinement by using COOT. Then the α -helices for the further polypeptides of the head module were manually built, and connected. Next, β -sheets were manually built into the unassigned structured regions in the electron density, which corresponded to the neck and fixed jaw domains. Subsequently, we began assigning the polypeptide identities at the neck domain. We tracked specific SeMet labelling patterns dictated by the presence of Met in the primary sequences of the polypeptides, and the presence of bulky regions corresponding to aromatic residue positions, as markers. We used secondary structure predictions for additional guidance. First, Med8 BH1, Med17 BH1 and Med22 BH2 were identified in α -helical bundle regions in the neck domain from their primary-sequence-specific, unique SeMet labelling pattern: these regions all contain more than two SeMet peaks and the spacing of SeMet peaks was consistent with the corresponding amino-acid sequences in the subunits. This assignment was consistent with the secondary structure predictions indicating α -helical structure. The remaining Med8 residues (60–170), as well as Med22 BH1, were assigned by tracing from the Med8 BH1 helix back to the Med8 C terminus, and by tracing from the Med22 BH2 helix back to the Med22 N terminus. This assignment was validated by the fact that their Met locations aligned with anomalous peaks on the experimental map. Next, we identified the NTD of Med6 based on its unique SeMet positions, and also identified Med6 BH based on a specific location of SeMet (Met48), a bulky aromatic ring (Phe52) (Supplementary Fig. 19a) and continuity from the NTD of Med6, consistent with secondary structure predictions. The C-terminal 80 residues of Med6 were proteolyzed in the crystals (Supplementary Fig. 1). Consequently, no density was found corresponding to the C terminus of Med6. We traced Med17 BH1, and identified the longest helix in the neck domain as the BH2 helix of Med17 on the basis of a single SeMet (Met313) and the aromatic side chain of Tyr269 (Supplementary Fig. 19b); this assignment was also consistent with secondary structure predictions. Finally, the two remaining continuous α -helices in the neck domain were identified as Med11 BH1 and Med11 BH2 because of one unique SeMet position of Med11. This assignment matches perfectly to the secondary structure prediction as well.

Next we focused on the fixed jaw domain. By subtracting the polypeptides already assigned to the neck and the movable jaw (see above), the fixed jaw should only contain the C-terminal regions of subunits Med11, Med17 and Med22. First, on the basis of continuity, SeMet position (Met422), aromatic ring position (Tyr423) (Supplementary Fig. 19c), and secondary structure prediction, we identified helix 420–455, β -sheet 456–480 and helices 496–523 and 540–570 of Med17. The remainder of the electron density in this region was continuous, and thus enabled us to trace Med17 completely to its C terminus. We identified the Med17 CTH and β -sheet with α -helix 600–608 on the basis of the SeMet positions and α -helix length from secondary structure prediction. Finally, we assigned two remaining helices: Med11 CTH was identified from the presence of one SeMet peak, and we assigned the last helix to Med22 CTH, which entirely lacks SeMet.

Initially, all models were refined using the program CNS DEN³⁵, refinement with strong NCS restraints between the three independent complexes in the asymmetric unit, and twinning refinement. Then the model was refined using PHENIX with NCS restraints and a single refined group isotropic temperature factor for each subunit, Ramachandran restraint, TLS refinement and twinning refinement. The geometry of the final model is good, with 91.3%, 8.0%, 0.7% of the amino-acid residues in the most favoured, allowed, and disallowed regions of the Ramachandran plot, respectively. All structural illustrations and electron density maps were prepared with PYMOL (<http://www.pymol.org/>) and COOT. PSIPRED was used for secondary structure prediction³⁶.

Docking of the X-ray structure into the electron microscopy map. The model of 12-subunit Pol II was docked into the Mediator–Pol II holoenzyme structure⁸ followed by docking of the X-ray model of the head module into the density corresponding to the Mediator head module, using the program CHIMERA³⁷.

Electron microscopy sample preparation, data collection and image analysis. We diluted purified head module deletion mutants in buffer containing 25 mM KCl, 25 mM Tris-HCl (pH 7.8) and 5 mM DTT. For preparation of all electron microscopy samples, about 3 µl of protein solution was applied to a carbon-coated Maxtaform, 300-mesh Cu/Rh EM specimen grid (Ted Pella) freshly glow-discharged in the presence of amylamine. The particles were then preserved by staining with a 2.0% (w/w) uranyl acetate solution using the sandwich carbon layer technique^{38,39}. The images were recorded under low-dose conditions using a Tecnai Spirit (Philips/FEI) microscope equipped with a LaB6 filament and operating at an accelerating voltage of 120 kV. Images were recorded on a Tietz (TVIPS) CCD camera at ×42,000 magnification and approximately 1-µm under-focus, resulting in a final pixel size corresponding to 5.06 Å.

The images were initially analysed using the ml_align2d program, a multi-reference, two-dimensional alignment routine with a maximum-likelihood target function⁴⁰ implemented in the XMIPP package⁴¹. Averages derived from the ml_align2d program were used to run iterative alternating rounds of supervised multi-reference alignment/classification and reference-free alignment as described previously⁴² to improve the homogeneity of the image classes.

In vitro transcription and the CTD phosphorylation assays. The *in vitro* transcription assay to assess activity of the recombinant head module and its mutant form using *srb4*⁴⁵ mutant crude extract was performed as described previously¹⁰. Quantification of transcripts on an absolute scale was performed using a FLA-5100 FUJIFILM fluorescent image analyser and the MultiGauge software package after

addition of 1 nCi of ³²P UTP to the gel 5 min before the end of the run. The CTD phosphorylation assay was performed as previously described¹⁵.

Yeast phenotypic analysis. The shuttle vectors carrying the MED17 mutations are described in Supplementary Table 4. The shuttle vectors were introduced into yeast strain Z572 by plasmid shuffling, and grown on SC medium containing 5-FOA at 30 °C as previously described¹⁵.

26. Li, M. Z. & Elledge, S. J. Harnessing homologous recombination *in vitro* to generate recombinant DNA via SLIC. *Nature Methods* **4**, 251–256 (2007).
27. Tempst, P., Geromanos, S., Elicone, C. & Erdjument-Bromage, H. Improvements in microsequencer performance for low picomole sequence analysis. *Methods* **6**, 248–261 (1994).
28. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
29. Adams, P. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
30. Sheldrick, G. A short history of SHELX. *Acta Crystallogr. A* **64**, 112–122 (2008).
31. McCoy, A. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
32. Cowtan, K. Recent developments in classical density modification. *Acta Crystallogr. D* **66**, 470–478 (2010).
33. Cowtan, K. Modified phased translation functions and their application to molecular-fragment location. *Acta Crystallogr. D* **54**, 750–756 (1998).
34. Emsley, P., Lohkamp, B., Scott, W. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
35. Schröder, G. F., Levitt, M. & Brunger, A. T. Super-resolution biomolecular crystallography with low-resolution data. *Nature* **464**, 1218–1222 (2010).
36. Bryson, K. *et al.* Protein structure prediction servers at University College London. *Nucleic Acids Res.* **33**, W36–W38 (2005).
37. Pettersen, E. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
38. Stoffer, G. & Stoffer-Meilicke, M. *The Ultrastructure of Macromolecular Complexes Studied with Antibodies* 409–455 (De Gruyter, 1983).
39. Tischendorf, G. W., Zeichhardt, H. & Stoffer, G. Determination of the location of proteins L14, L17, L18, L19, L22, L23 on the surface of the 50S ribosomal subunit of *Escherichia coli* by immune electron microscopy. *Mol. Gen. Genet.* **134**, 187–208 (1974).
40. Scheres, S. H. *et al.* Maximum-likelihood multi-reference refinement for electron microscopy images. *J. Mol. Biol.* **348**, 139–149 (2005).
41. Sorzano, C. O. *et al.* XMIPP: a new generation of an open-source image processing package for electron microscopy. *J. Struct. Biol.* **148**, 194–204 (2004).
42. Brignole, E. J., Smith, S. & Asturias, F. J. Conformational flexibility of metazoan fatty acid synthase enables catalysis. *Nature Struct. Mol. Biol.* **16**, 190–197 (2009).

Initial radiation of jaws demonstrated stability despite faunal and environmental change

Philip S. L. Anderson¹, Matt Friedman², Martin D. Brazeau^{3†} & Emily J. Rayfield¹

More than 99 per cent of the roughly 58,000 living vertebrate species have jaws¹. This major clade, whose members are collectively known as gnathostomes ('jawed mouths'), made its earliest definitive appearance in the Silurian period, 444–416 million years (Myr) ago, with both the origin of the modern (crown-group) radiation and the presumptive invasion of land occurring by the end of the Devonian period² (359 Myr ago). These events coincided with a major faunal shift that remains apparent today: the transition from Silurian ecosystems dominated by jawless fishes (agnathans) to younger assemblages composed almost exclusively of gnathostomes^{2,3}. This pattern has inspired several qualitative descriptions of the trophic radiation and ecological ascendance of the earliest jawed vertebrates^{3–7}. Here we present a quantitative analysis of functional variation in early gnathostome mandibular elements, placing constraints on our understanding of evolutionary patterns during this critical interval. We document an initial increase in functional disparity in the Silurian that stabilized by the first stage of the Devonian, before the occurrence of an Emsian (~400 Myr ago) oxygenation event implicated in the trophic radiation of vertebrates⁸. Subsequent taxonomic diversification during the Devonian did not result in increased functional variation; instead, new taxa revisited and elaborated on established mandibular designs. Devonian functional space is dominated by lobe-finned fishes and 'placoderms'; high disparity within the latter implies considerable trophic innovation among jaw-bearing stem gnathostomes. By contrast, the major groups of living vertebrates—ray-finned fishes and tetrapods—show surprisingly conservative mandibular morphologies with little indication of functional diversification or innovation. Devonian gnathostomes reached a point where they ceased to accrue further mandibular functional disparity before becoming taxonomic dominants relative to 'ostracoderm'-grade jawless fishes, providing a new perspective on classic adaptive hypotheses concerning this fundamental shift in vertebrate biodiversity.

Morphological disparity has often been used as a proxy for ecological/functional variety in the fossil record⁹. However, divergent anatomies can be biomechanically similar^{10–12}, leading to decoupling of morphological and functional metrics^{12–14}. To evaluate potential functional diversity, we have devised a measure of variation derived from biomechanical traits. We focus on the mandible because it is well characterized functionally in modern fishes¹⁵ and because it is the key character proposed by adaptive schemes to separate the gnathostomes from their jawless predecessors^{4,6}. We document how functional variation accumulates after the origin of jaws and test the hypothesis that, during their rise to dominance, gnathostomes outcompeted or displaced jawless fishes.

We assembled a data set consisting of 198 Silurian/Devonian gnathostome genera scored for 31 mandibular traits (11 continuous and 20 discrete), many of which have been shown to be functionally relevant in living taxa. These data were analysed using non-metric multidimensional scaling, yielding a multivariate 'function space'. Functional disparity was calculated across eight time bins (the late

Silurian plus the seven stages of the Devonian) using a variety of measures (sum of variances is presented here because it is relatively robust to variation in sample size¹⁶). We tested for differences between successive bins using multiple approaches (likelihood ratios and *t*-tests), and used a jackknifing procedure that pruned exceptional fossil assemblages from our data set to examine the sensitivity of results to 'Lagerstätten effects', biases arising from localities characterized by exceptional preservation or fossil abundance. Full details of our methodology are given in Supplementary Information.

Our results show remarkably stable levels of lower-jaw functional disparity during the Devonian (Fig. 1a). There is a trend of increasing disparity from the late Silurian to the end of the Early Devonian (the Emsian stage); however, by the earliest Devonian (the Lockhovian stage, ~415 Myr ago) functional mandibular disparity attained a level statistically indistinguishable from that found in the Late Devonian. When *P* values are adjusted for multiple comparisons, the only significant disparity shift between successive Devonian stages (corrected *P* value, 0.013) is a decline on entering the Middle Devonian (the Eifelian stage, ~398 Myr ago; Fig. 1a). This seems to reflect a reduction in the number of marine sites sampled in the Eifelian, combined with the Emsian Taemas/Wee Jasper fauna, which contains a series of extensively studied durophagous lungfishes with exceptionally robust jaws. When this site is removed in jackknife analyses, the drop in disparity is no longer significant (corrected *P* value, 0.112). When differences are assessed using likelihood ratios, none exceeds a threshold value of 8 (ref. 17; likelihood ratio range, 1.023–4.34). This pattern of functional stability arises despite an increase both in the number of mandibles sampled per bin over our study interval and in the taxonomic diversity of gnathostomes during the Devonian¹⁸, and is robust to the removal of exceptional Lagerstätten. Details of all analyses are given in Supplementary Information.

A breakdown of morphospace by stage shows that initial functional disparity in the late Silurian was localized in the lower-left and upper-right quadrants (Fig. 2). The former is populated by gnathostomes (for example stem sarcopterygians and ischnacanthid 'acanthodians') with gracile mandibles bearing slender, pointed cusps and characterized by low closing mechanical advantages, whereas the latter contains taxa (for example mesacanthid acanthodians) with toothless jaws. The Lockhovian expansion into the lower-right quadrant reflects the appearance of taxa (for example the lungfish *Diabolepis*) with robust jaws characterized by both high closing mechanical advantage and high second moment of area, two features consistent with durophagy. Subsequent mandibular evolution during the Devonian resulted in novel designs that elaborated existing ones, but did not substantially exceed the boundaries established earlier in the interval (Fig. 2). Specific anatomical correlates of our ordination axes are provided in Supplementary Table 2.

We examined patterns of partial disparity¹⁹ to dissect the contributions to overall functional variation made by major taxonomic assemblages: 'Acanthodii', an extinct assemblage sometimes called spiny sharks; Actinopterygii, or ray-finned fishes, related to most fishes alive today; Chondrichthyes, or cartilaginous fishes, the precursors of rays, sharks and ratfishes; 'Placodermi', armoured stem gnathostomes; and

¹Department of Earth Sciences, University of Bristol, Bristol BS8 1RJ, UK. ²Department of Earth Sciences, University of Oxford, Oxford OX1 3AN, UK. ³Museum für Naturkunde, Leibniz Institute for Research on Evolution and Biodiversity at the Humboldt University in Berlin, 10115 Berlin, Germany. [†]Present address: NCB Naturalis, 2300 RA Leiden, The Netherlands.

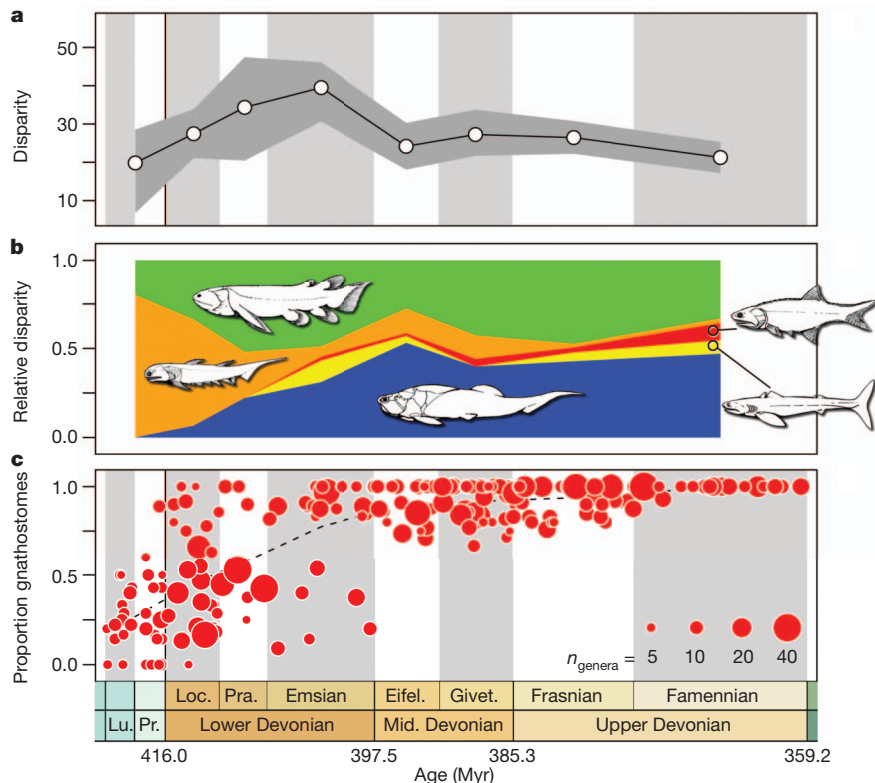


Figure 1 | Functional mandibular disparity among Silurian/Devonian gnathostomes. All horizontal axes show time as indicated at bottom.

a, Disparity (sum of variances) across eight time bins. The dark-grey region spans the 95% confidence intervals based on 1,000 bootstrap pseudoreplicates. **b**, Relative contributions (partial disparity) of major gnathostome groups to overall functional disparity. Orange, 'Acanthodii'; green, Sarcopterygii; blue, 'Placodermi'; yellow, Chondrichthyes; red, Actinopterygii. **c**, Faunal

composition data for the late Silurian and Devonian. Discs represent individual vertebrate assemblages plotted as a function of time and proportion of gnathostomes that comprise those faunas (discs jittered within time bins for clarity). The area of each disc is proportional to the total number of vertebrate genera represented, n_{genera} . Eifel., Eifelian stage; Givet., Givetian stage; Loc., Lockhovian stage; Lu., Ludlow series; Pr., Pridoli series; Pra., Pragian stage.

Sarcopterygii, or lobe-finned fishes, including lungfishes and terrestrial vertebrates. Sarcopterygians make a major and relatively stable contribution to mandibular disparity throughout our study interval (Fig. 1b). The earliest definitive digit-bearing tetrapods (*Acanthostega* and *Ichthyostega*) and their immediate relatives (*Ventastega*, *Elginerpeton*, *Densignathus* and *Metaxygnathus*) show a conservative pattern in mandibular function relative to sarcopterygian 'fishes' (Fig. 2). This is in contrast to studies that report a distinct shift in cranial suture patterns between aquatic taxa and early digit-bearing tetrapods²⁰, suggesting that different parts of the feeding system evolved in separate phases. By contrast, lungfishes are a major source of mandibular innovation among sarcopterygians, colonizing a new region of functional space in the Lockhovian and occupying it exclusively throughout the Devonian (Fig. 2).

'Placoderms' occupy 50% of realized Devonian gnathostome morphospace volume (Fig. 2), indicating that considerable functional differentiation occurred on the gnathostome stem, independently of the origin of modern jawed vertebrates. 'Placoderm' functional variety remains undiminished until their sudden extinction at the Devonian/Carboniferous boundary. A direct reading of our results implies that 'placoderms' were a major contributor to gnathostome disparity only from the Middle Devonian onward. We argue that this is an artefact; many 'placoderms' are known from the Early Devonian and, increasingly, the Silurian^{2,21}, but few preserve mandibular material that can be included in our analysis. This suggests that disparity measures for the Silurian and Early Devonian are probably conservative underestimates, and that the addition of more 'placoderms' of this age would only reinforce the pattern of early functional morphospace saturation.

The remaining three gnathostome assemblages—'acanthodians', chondrichthyans and actinopterygians—make relatively minor

contributions to overall disparity (Fig. 1b). 'Acanthodians' make a major contribution to functional variety during the Silurian, but their relative importance wanes substantially throughout the Devonian, reflecting a decline in functional diversity and apparent richness combined with the taxonomic ascendance of other gnathostome clades²². Actinopterygians, the most speciose group of living jawed fishes¹, contribute little to overall disparity in the Devonian, with their few representatives conservatively clustering within tightly delimited regions of function space (Fig. 2). This is consistent with previous arguments, drawn from taxonomic data, that actinopterygians were not ecologically diverse until the Carboniferous period²². The reliability of low measures of chondrichthyan partial disparity is unclear, however, because the numerous Devonian chondrichthyans known exclusively from dental remains²³ cannot be included in our analysis.

Our results provide a new perspective on debates concerning the Devonian shift from agnathan- to gnathostome-dominated fossil assemblages, a transition that remains apparent in modern vertebrate fauna^{1,22}. A range of schemes, largely derived from anecdotal evidence, have sprung up in response to this pattern of turnover^{3–7}. Faunal data clearly show that gnathostomes shared habitat space evenly with 'ostracoderm'-grade agnathans well into the Early Devonian (Fig. 1c). It is only during and after the Emsian that gnathostomes became taxonomic dominants in most fossil assemblages, several million years after they explored the extremes of realized Devonian mandibular function space. The offset between the trophic radiation of gnathostomes (Fig. 1a) and the taxonomic collapse of 'ostracoderms' (Fig. 1c), combined with the long history of coexistence between these two groups, argues against the direct ecological replacement of jawless fishes by jawed forms⁶. Furthermore, we find no evidence for a significant positive relationship between mandibular disparity within assemblages and the taxonomic

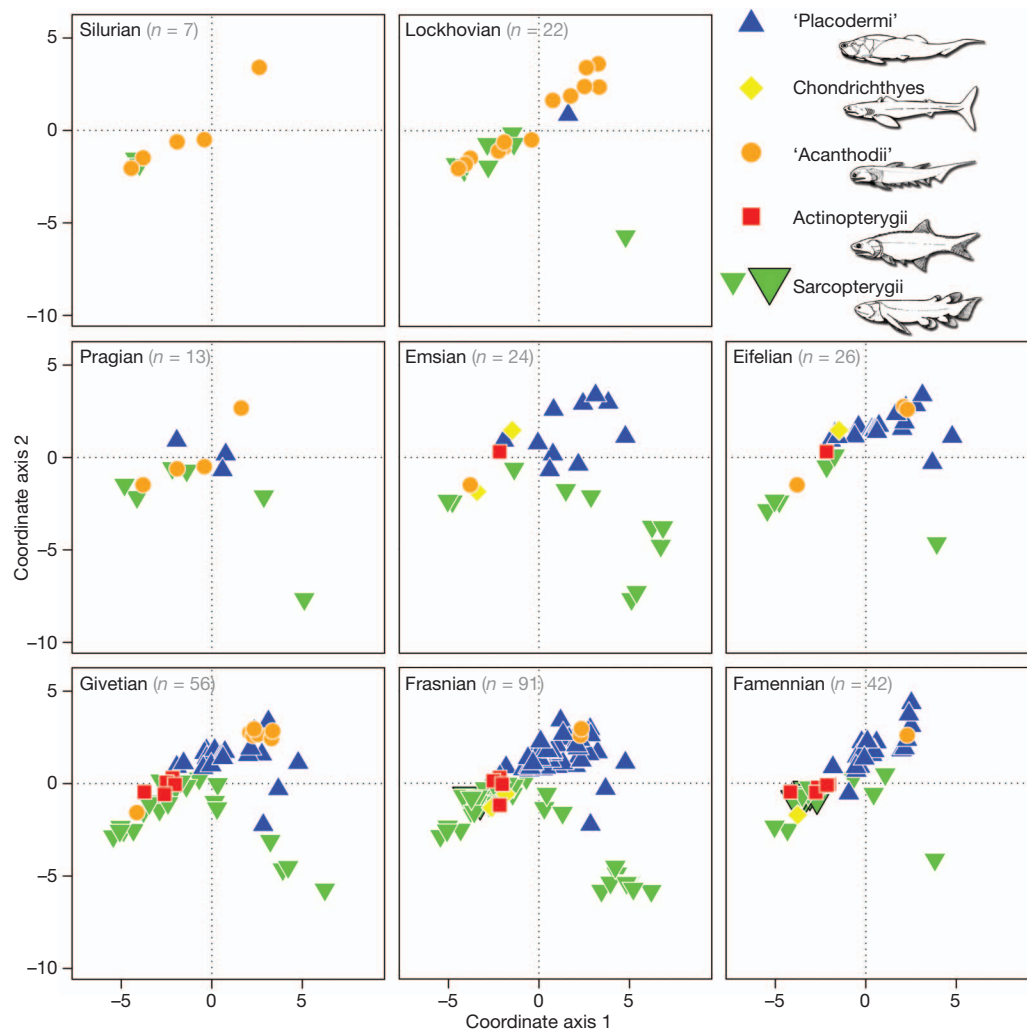


Figure 2 | Patterns of functional morphospace occupation for jawed vertebrates during the late Silurian and Devonian. Points represent averages for genera plotted on coordinate axes 1 and 2 of the non-metric multidimensional scaling ordination. Orange circles, 'Acanthodii'; green down-triangles, Sarcopterygii (tetrapods indicated by larger symbols outlined in

black); blue up-triangles, 'Placodermi'; yellow diamonds, Chondrichthyes; red squares, Actinopterygii. The area in the lower-right corner of the functional morphospace is occupied solely by lungfishes. See text and Supplementary Information for further details.

dominance of jawed vertebrates within those faunas (measured as the proportion of gnathostomes; Pearson and Kendall rank-order correlations, $P = 0.90\text{--}0.94$; weighted least-squares regression, $P = 0.40\text{--}0.44$). This pattern is inconsistent with the assertion⁶ that gnathostome diversification primarily reflects the ecological displacement of agnathans and subsequent refilling of their functional roles. Indeed, few of the feeding strategies apparently taken up by early gnathostomes, such as macropredation and durophagy, have any clear analogue among jawless taxa³. Instead, the proliferation of gnathostomes probably reflects new ecological opportunities associated with the origin of jaws and, perhaps just as importantly, the shift from benthic lifestyles to pelagic ones²⁴.

The Devonian radiation of gnathostomes is marked by an early plateau of functional disparity and a consistent range of mandibular variation in the face of major environmental perturbations, including fluctuating global temperatures²⁵, shifting continental weathering patterns²⁶, sea level changes²⁶ and a major extinction event²⁷. Various proposals have linked patterns of early vertebrate evolution to changes in atmospheric composition, with parallel claims that gnathostome diversification was driven by increases in oxygen levels during the Silurian²⁸ or the Devonian⁸. Silurian data are too scant to comment on the first proposal, but we note that gnathostomes achieved stable levels of functional diversity before the occurrence of an Emsian oxygenation event (~400 Myr ago) implicated in the trophic radiation

of vertebrates⁸. There is no significant shift in functional disparity between the Frasnian and Famennian stages (~375 Myr ago; Fig. 1) correlating with the Kellwasser event, which is classically identified as one of the 'Big Five' Phanerozoic extinctions²⁷. This event resulted in a global reef collapse²⁹; however, jawed vertebrates seem to be unaffected in terms of diversity²², functional disparity and potentially ecological variation. More broadly, our results imply constraints on the functional diversity of the gnathostome mandibular system, with early functional saturation and subsequent refilling of established biomechanical roles by new taxonomic actors.

Received 16 March; accepted 18 May 2011.

Published online 6 July 2011.

- Nelson, J. S. *Fishes of the World* 4th edn (Wiley, 2006).
- Janvier, P. *Early Vertebrates* (Clarendon, 1996).
- Janvier, P. & Blicek, A. in *Palaeozoic Vertebrate Biostratigraphy and Biogeography* (ed. Long, J. A.) 67–86 (Johns Hopkins Univ. Press, 1993).
- Halstead, L. B. *The Pattern of Vertebrate Evolution* 26 (Oliver & Boyd, 1969).
- Colbert, E. H. *Evolution of the Vertebrates* 28 (Wiley, 1991).
- Carroll, R. L. *Vertebrate Paleontology and Evolution* 44 (Freeman, 1988).
- Pough, F. H., Janis, C. M. & Heiser, J. B. *Vertebrate Life* 8th edn (Pearson Benjamin Cummings, 2009).
- Dahl, T. W. *et al.* Devonian rise in atmospheric oxygen correlated to the radiations of terrestrial plants and large predatory fish. *Proc. Natl Acad. Sci. USA* **107**, 17911–17915 (2010).
- Foote, M. Morphological disparity in Ordovician–Devonian crinoids and the early saturation of morphological space. *Paleobiology* **20**, 320–344 (1994).

10. Alfaro, M. E., Bolnick, D. I. & Wainwright, P. C. Evolutionary dynamics of complex biomechanical systems: an example using the four-bar mechanism. *Evolution* **58**, 495–503 (2004).
11. Alfaro, M. E., Bolnick, D. I. & Wainwright, P. C. Evolutionary consequences of many-to-one mapping of jaw morphology to mechanics in labrid fishes. *Am. Nat.* **165**, E140–E154 (2005).
12. Wainwright, P. C. Functional versus morphological diversity in macroevolution. *Annu. Rev. Ecol. Evol. Syst.* **38**, 381–401 (2007).
13. Hulsey, C. D. & Wainwright, P. C. Projecting mechanics into morphospace: disparity in the feeding system of labrid fishes. *Proc. R. Soc. Lond. B* **269**, 317–326 (2002).
14. Anderson, P. S. L. Biomechanics, functional patterns, and disparity in Late Devonian arthrodires. *Paleobiology* **35**, 321–342 (2009).
15. Westneat, M. W. Evolution of levers and linkages in the feeding mechanisms of fishes. *Integr. Comp. Biol.* **44**, 378–389 (2004).
16. Ciampaglio, C. N., Kemp, M. & McShea, D. W. Detecting changes in morphospace occupation patterns in the fossil record: characterization and analysis of measures of disparity. *Paleobiology* **27**, 695–715 (2001).
17. Royall, R. M. *Statistical Evidence: A Likelihood Paradigm* (Chapman & Hall, 1997).
18. Benton, M. J. *The Fossil Record 2* (Chapman & Hall, 1993).
19. Foote, M. Contributions of individual taxa to overall morphological disparity. *Paleobiology* **19**, 403–419 (1993).
20. Markey, M. J. & Marshall, C. R. Terrestrial-style feeding in a very early aquatic tetrapod is supported by evidence from experimental analysis of suture morphology. *Proc. Natl Acad. Sci. USA* **104**, 7134–7138 (2007).
21. Zhao, W.-J. & Zhu, M. Siluro-Devonian vertebrate biostratigraphy and biogeography of China. *Palaeoworld* **19**, 4–26 (2010).
22. Sallan, L. C. & Coates, M. I. End-Devonian extinction and a bottleneck in the early evolution of modern jawed vertebrates. *Proc. Natl Acad. Sci. USA* **99**, 8139–8144 (2002).
23. Ginter, M., Hampe, O. & Duffin, C. J. *Handbook of Paleichthyology Vol. 3D: Chondrichthyes Paleozoic Elasmobranchii* (Pfeil, 2010).
24. Klug, C. *et al.* The Devonian nekton revolution. *Lethaia* **43**, 465–477 (2010).
25. Joachimski, M. M. *et al.* Devonian climate and reef evolution: insights from oxygen isotopes in apatite. *Earth Planet. Sci. Lett.* **284**, 599–609 (2009).
26. van Geldern, R. *et al.* Carbon, oxygen and strontium isotope records of Devonian brachiopod shell calcite. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **240**, 47–67 (2006).
27. Raup, D. M. & Sepkoski, J. J. Jr. Mass extinctions in the marine fossil record. *Science*, **215**, 1501–1503 (1982).
28. Qu, Q. M., Zhu, M. & Zhao, W. J. Silurian atmospheric O₂ changes and the early radiation of gnathostomes. *Palaeoworld* **19**, 146–159 (2010).
29. Copper, P. Reef development at the Frasnian/Famennian mass extinction boundary. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **181**, 27–65 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank P. Donoghue, M. Rücklin and M. Ruta for discussions. This work was supported by Royal Society and Marie-Curie Actions fellowships, awarded to P.S.L.A.; a Fell Fund award to M.F.; NERC grant NE/G016623/1, awarded in part to E.J.R.; and a QFRNT postdoctoral fellowship, to M.D.B.

Author Contributions P.S.L.A. designed and led the study, developed the biomechanical traits, collected data, performed the disparity and multivariate analyses, and wrote the paper. M.F. gathered data, wrote analytical code, performed the faunal analyses, drafted figures and wrote the paper. M.D.B. gathered data and produced illustrations. E.J.R. gave advice on biomechanical theory. All authors contributed to interpretation of the results and edited the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to P.S.L.A. (phil.anderson@bristol.ac.uk).

Structure and mechanism of the Swi2/Snf2 remodeller Mot1 in complex with its substrate TBP

Petra Wollmann^{1,2}, Sheng Cui^{1†}, Ramya Viswanathan³, Otto Berninghausen¹, Melissa N. Wells³, Manuela Moldt¹, Gregor Witte^{1,2,4}, Agata Butryn¹, Petra Wendler¹, Roland Beckmann^{1,4}, David T. Auble³ & Karl-Peter Hopfner^{1,2,4}

Swi2/Snf2-type ATPases regulate genome-associated processes such as transcription, replication and repair by catalysing the disruption, assembly or remodelling of nucleosomes or other protein–DNA complexes^{1,2}. It has been suggested that ATP-driven motor activity along DNA disrupts target protein–DNA interactions in the remodelling reaction^{3–5}. However, the complex and highly specific remodelling reactions are poorly understood, mostly because of a lack of high-resolution structural information about how remodellers bind to their substrate proteins. Mot1 (modifier of transcription 1 in *Saccharomyces cerevisiae*, denoted BTA1 in humans) is a Swi2/Snf2 enzyme that specifically displaces the TATA box binding protein (TBP) from the promoter DNA and regulates transcription globally by generating a highly dynamic TBP pool in the cell^{6,7}. As a Swi2/Snf2 enzyme that functions as a single polypeptide and interacts with a relatively simple substrate, Mot1 offers an ideal system from which to gain a better understanding of this important enzyme family. To reveal how Mot1 specifically disrupts TBP–DNA complexes, we combined crystal and electron microscopy structures of Mot1–TBP from *Encephalitozoon cuniculi* with biochemical studies. Here we show that Mot1 wraps around TBP and seems to act like a bottle opener: a spring-like array of 16 HEAT (huntingtin, elongation factor 3, protein phosphatase 2A and lipid kinase TOR) repeats grips the DNA-distal side of TBP via loop insertions, and the Swi2/Snf2 domain binds to upstream DNA, positioned to weaken the TBP–DNA interaction by DNA translocation. A ‘latch’ subsequently blocks the DNA-binding groove of TBP, acting as a chaperone to prevent DNA re-association and ensure efficient promoter clearance. This work shows how a remodelling enzyme can combine both motor and chaperone activities to achieve functional specificity using a conserved Swi2/Snf2 translocase.

Mot1 is highly conserved among eukaryotes and consists of an amino-terminal TBP binding region of approximately 90–140 kDa with predicted HEAT repeats, followed by a carboxy-terminal Swi2/Snf2-type ATPase domain of approximately 60–70 kDa (refs 8, 9). To provide a structural framework for a remodeller–substrate complex, we determined the crystal structure of the N-terminal domain (NTD) of *Encephalitozoon cuniculi* (Ec) Mot1 (comprising the HEAT domain, residues 1–779, but lacking the ATPase domain, residues 780–1256) in complex with full-length EcTBP, to 3.1 Å resolution (Fig. 1 and Supplementary Table 1). EcMot1 has the characteristic sequence and biochemical features of *S. cerevisiae* (Sc) Mot1 and human BTA1, including TBP- and DNA-stimulated ATPase activity, TBP binding via its HEAT domain, and, most importantly, ATP-stimulated TBP displacement from TATA DNA (Supplementary Figs 1 and 2).

The EcMot1 NTD consists of a highly elongated stretch of 16 HEAT repeats, arranged in a horseshoe shape with dimensions of about 95 Å × 85 Å × 40 Å, and it forms a specific 1:1 complex with EcTBP (Fig. 1). Notably, Mot1 wraps around one side of the pseudosymmetric

TBP and grips both the convex protein-interacting surface and the concave DNA-binding surface of TBP via several loop insertions in the array of HEAT repeats. This wrapping interaction enables Mot1 to split the very stable EcTBP dimer that forms in the absence of DNA¹⁰, and that we observed biochemically and in a separate crystal structure of EcTBP alone at 1.9 Å resolution (Supplementary Fig. 3a and Supplementary Table 1). Despite this dual-sided grip, Mot1 does not alter the structure of TBP substantially because EcTBP bound to Mot1, EcTBP in the TBP dimer and ScTBP bound to DNA are all very similar (Supplementary Fig. 3c–f). This indicates that remodelling of TBP does not proceed via changes in TBP structure as a simple consequence of Mot1 binding, but requires the ATP-dependent action of the Swi2/Snf2 domain.

Promoter-bound TBP has its DNA-binding surface occupied, so Mot1 uses highly complementary HEAT-repeat loops to recognize the convex protein-interaction surface of TBP (Fig. 2a). In TBP, α-helices H1 and H2 are bound by the loop of HEAT repeat 4 (residues 209–221), and by interactions with α-helix 13 in HEAT repeat 5 and α-helix 15 in HEAT repeat 6. Most of these interactions are ion pairs between TBP R46 and Mot1 Q256, TBP R48 and Mot1 D212, TBP R65 and Mot1 D215, TBP R96 and Mot1 D216, TBP K99 and Mot1 D216, TBP K103 and Mot1 D290 and TBP K103 and Mot1 D292 (Supplementary Table 2). In addition, Mot1 F213 binds to a hydrophobic cleft between H1, H2 and β-sheet S2, providing a hydrophobic anchor, and Mot1 residues F210 and W255 pack against the side chains of TBP residues R48 and K103.

These interactions are well conserved evolutionarily (Supplementary Fig. 4a and Supplementary Table 2) and are supported by functional data *in vivo* and *in vitro*. For instance, ScTBP K145 (EcTBP K103) is an essential residue for stabilization of the ScMot1–ScTBP interaction⁸. We mutated K103 in EcTBP and observed that EcTBP(K103E) failed to form a stable complex with EcMot1(NTD) *in vitro* (Fig. 2b). Moreover, mutation of D365 (D212 in EcMot1) inactivated ScMot1 *in vivo* and abolished the Mot1–TBP interaction *in vitro*⁸. Mutations of K138 in ScTBP also impaired the interaction with ScMot1, consistent with the projection of the homologous side chains into the EcMot1(NTD)–EcTBP interface^{8,11}. The distribution of residues along the length of the EcMot1 NTD is also consistent with earlier work showing that broad segments of the ScMot1 and BTA1 N termini are important for stable interaction with TBP^{8,9,12}. Thus, the specific interaction interface between the Mot1 HEAT repeats and the convex surface of TBP is well suited to provide specific recognition of the TBP surface in the TBP–promoter complex, explaining why Mot1 specifically targets TBP–DNA and not other protein–DNA complexes.

Unexpectedly, the concave DNA-binding surface of TBP, accessible only when TBP is displaced from promoter DNA, is bound by Mot1 as well (Fig. 2c). A long ‘latch’, located between HEAT repeats 2 and 3, protrudes from the side of Mot1 distal to TBP and wraps all the way

¹Department of Biochemistry, Ludwig-Maximilians-University, Feodor-Lynen-Strasse 25, 81377 Munich, Germany. ²Munich Center for Advanced Photonics, Gene Center Munich, Feodor-Lynen-Strasse 25, 81377 Munich, Germany. ³Department of Biochemistry and Molecular Genetics, University of Virginia Health System, Charlottesville, Virginia 22908, USA. ⁴Center for Integrated Protein Science, Gene Center Munich, Feodor-Lynen-Strasse 25, 81377 Munich, Germany. †Present address: Institute of Pathogen Biology, Chinese Academy of Medical Sciences and Peking Union Medical College, No. 9, Dong Dan San Tiao, Beijing 100730, China.

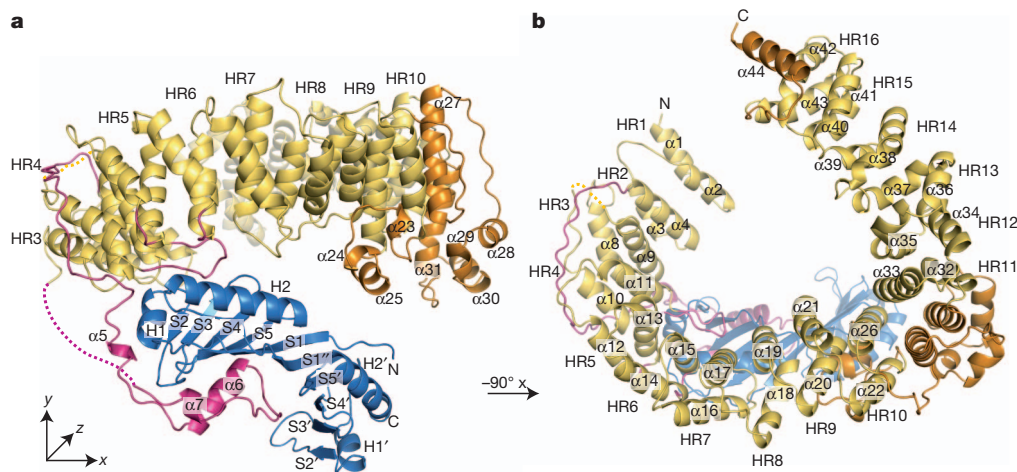


Figure 1 | Overview of the EcMot1(NTD)-EcTBP structure. **a, b,** Structure of the EcMot1(NTD)-EcTBP complex in ribbon representation with highlighted and annotated secondary structure. The HEAT repeats (HR) of EcMot1(NTD) are coloured yellow and non-HEAT-repeat insertions are in

orange. The latch and the loops of HR 4 to HR 6 are highlighted in magenta. EcTBP is coloured blue. Two loops not traced by electron density are indicated by dashed lines.

around the side of Mot1 and TBP. Notably, its tip (residues 101–130) substitutes for interactions made by four base pairs (bp) at and immediately downstream from the TATA sequence (Fig. 2d). A set of hydrophobic interactions matches the hydrophobic nature of TBP's DNA-binding groove. For instance, the side chain of M109 in Mot1 replaces a deoxyribose moiety in binding to TBP F57, a prominent and highly conserved DNA-binding residue of TBP. The main chain of residues 118–129 folds along the position of the backbone of the coding DNA strand, with side chains often placed at positions occupied by base and sugar moieties. F123 in Mot1 replaces a deoxyribose moiety

and stacks with the conserved TBP Q116, and F129 in Mot1 replaces a base moiety that interacts with the aromatic pair F57 and F74 in TBP.

To test the function of the latch, we generated the mutants EcMot1(Δ latch) and EcMot1(NTD Δ latch) that lack residues 96–132. Both proteins can still interact with EcTBP with approximately equal Mot1:TBP molarity (Supplementary Fig. 4b). This observation indicates that EcTBP is mainly bound by acidic loops of HEAT repeats 4–6 in Mot1. However, the latch might prevent TBP rebinding to DNA (after DNA dissociation) and might also prevent homodimerization by saturating the exposed, hydrophobic DNA-binding cleft of TBP (see

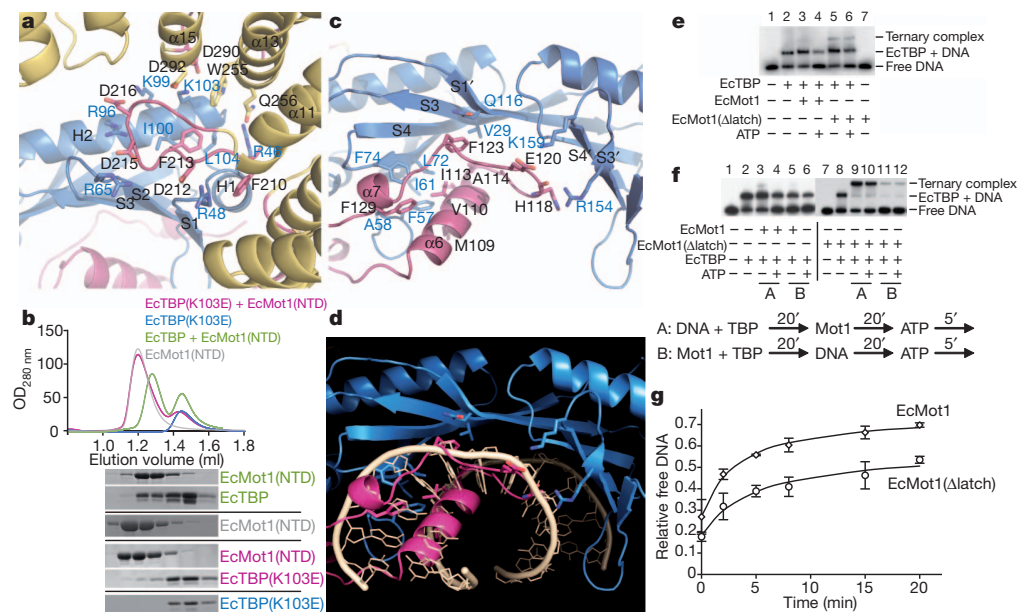


Figure 2 | Details of the interaction interfaces and latch function. **a,** Close-up view of the EcMot1-EcTBP interaction (colour scheme as in Fig. 1). **b,** Wild-type EcTBP and EcMot1(NTD) (green) can form a stable complex, whereas the EcTBP(K103E) mutant does not co-elute with EcMot1(NTD) (pink) in size exclusion chromatography (Supplementary Fig. 1b). **c, d,** The latch of EcMot1 (pink, shown in **c**) overlaps with the DNA-binding region (shown in **d**) of EcTBP (blue). Some bases of the superimposed DNA (wheat, from PDB 1YTB²⁹) were omitted. **e, f,** Electrophoretic mobility shift assays (for corresponding quantifications see Supplementary Fig. 4). **e,** EcMot1(Δ latch) formed stable ternary complexes with EcTBP-DNA (lane 5). However,

although wild-type EcMot1 largely cleared the DNA probe of bound TBP in an ATP-dependent reaction (lane 4), EcMot1(Δ latch) was less efficient in TBP removal (lane 6). **f,** EcMot1 was incubated with EcTBP after (A) or before (B) the addition of DNA. Preincubation of the two proteins inhibited TBP's ability to bind DNA. **g,** EcMot1(Δ latch) dissociated EcTBP-DNA less efficiently than wild-type EcMot1. ATP was added to pre-formed EcMot1-EcTBP-DNA or EcMot1(Δ latch)-EcTBP-DNA ternary complexes, and the proportion of free DNA was quantified by electrophoretic mobility shift assays at various times thereafter. Data represent mean and standard error from two independent experiments.

Supplementary Fig. 3b). Indeed, whereas EcMot1(NTD) forms a heterodimer with EcTBP, we found that EcMot1(NTD Δ latch) forms a 2:2 complex with EcTBP (Supplementary Table 3). The most likely explanation is that two EcMot1(NTD Δ latch) molecules bind the EcTBP dimer, but fail to dissociate the dimer owing to the absence of the latch. Because EcMot1(Δ latch) in complex with EcTBP does not show a substantially increased hydrodynamic radius compared to the wild-type complex in gel filtration (Supplementary Fig. 4b), it is likely that the Swi2/Snf2 domain sterically prevents dimerization of EcMot1(Δ latch) via TBP dimers.

Thus, although one function of the latch might be to keep TBP in a monomeric state, a more intriguing role might be to interfere with DNA binding by TBP. To test this, we analysed the ability of the EcMot1(Δ latch) protein to bind to the TBP–DNA complex. In contrast to wild-type EcMot1, EcMot1(Δ latch) formed readily detectable ternary complexes with EcTBP and DNA (Fig. 2e, f), indicating that the latch makes the association of EcMot1 with EcTBP–DNA less stable. Although it bound to TBP–DNA more efficiently, EcMot1(Δ latch) was notably impaired in ATP-dependent TBP–DNA dissociation (Fig. 2e–g and Supplementary Fig. 4d, e). This was not due to a defect in ATPase activity (Supplementary Fig. 4g). Moreover, when combined with EcTBP before DNA addition, EcMot1 inhibited DNA binding by EcTBP (Fig. 2f and Supplementary Fig. 4e). EcMot1(NTD) also inhibited DNA binding by EcTBP in a reaction that required the latch (Supplementary Fig. 4c, f). However, the latch was not essential for inhibiting the EcTBP–DNA interaction in the context of the full-length EcMot1 protein (Fig. 2f and Supplementary Fig. 4e), indicating that

both the latch and the ATPase domain can modulate EcTBP DNA-binding activity. Taken together, the data indicate that the latch has ‘chaperone’ activity and regulates macromolecular interactions with the hydrophobic groove of TBP. Because DNA binding and latch binding to TBP are mutually exclusive (Fig. 2d), it is unlikely that the latch initially disrupts the TBP–DNA complex. Consistent with this, EcMot1(Δ latch) was able to displace TBP from DNA using ATP, but the overall level of displacement was increased by the latch (Fig. 2g). Thus, our combined data can be explained by a physiologically plausible model in which the ATP-dependent action of the Swi2/Snf2 domain remodels TBP–TATA first, and then the latch blocks the exposed hydrophobic groove to prevent rebinding.

To reveal the architecture of the whole *E. cuniculi* Mot1–TBP complex, including its Swi2/Snf2 domain, we generated three-dimensional reconstructions of negatively stained EcMot1–EcTBP particles visualized in electron micrographs (Fig. 3a and Supplementary Fig. 5). The three-dimensional reconstruction is shaped like a slightly closed ‘C’ with a globular protrusion, and is similar to the three-dimensional reconstructions of the human TBP–BTAF1 complex¹³. To locate the Swi2/Snf2 domain unambiguously, we imaged a complex of EcTBP with a deletion mutant of EcMot1 in which the C-terminal half of the Swi2/Snf2 domain was truncated (EcMot1(Δ CT)) (Supplementary Figs 5 and 6c). We found that the prominent protrusion is missing from this complex, indicating that this protrusion corresponds to the C-terminal half of the ATPase (Fig. 3b). Finally, we imaged Mot1 without TBP (Supplementary Figs 5 and 6b). Although Mot1 alone is evidently more flexible than it is in the Mot1–TBP complex, and

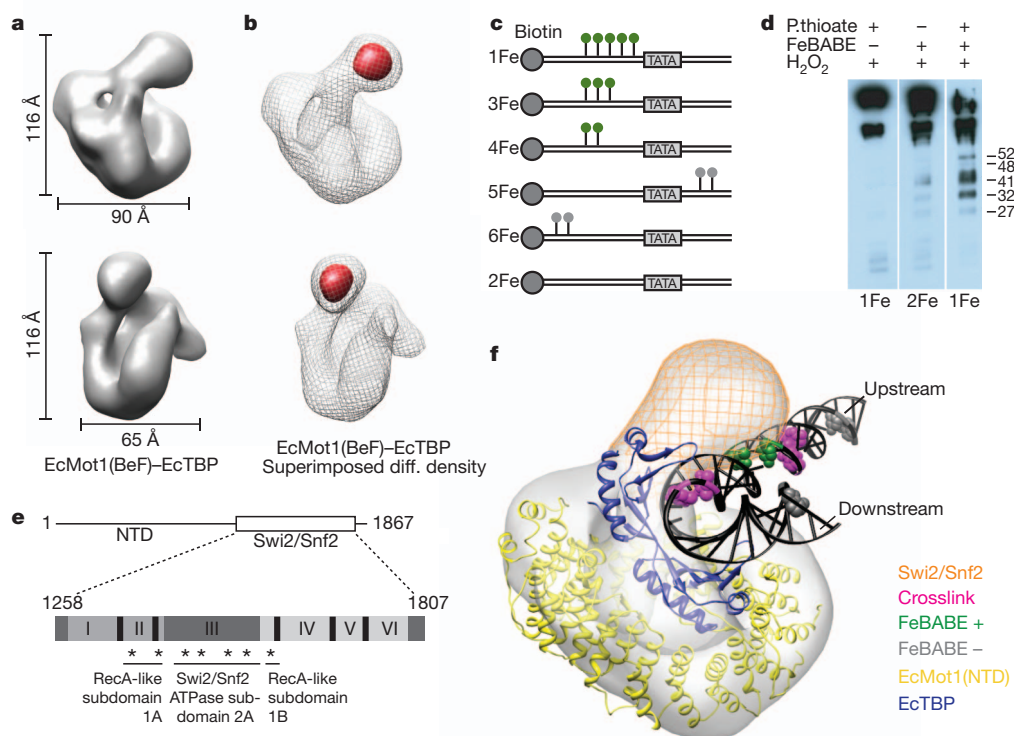


Figure 3 | Three-dimensional reconstruction of the EcMot1–EcTBP complex and model of the EcMot1–EcTBP–DNA complex. **a**, Two views of the EcMot1(BeF)–EcTBP density map. ADP–BeF₃[–] was added owing to its assumed stabilization of the ATPase domain. **b**, Subtraction map (red) between EcMot1(BeF)–EcTBP (grey mesh) and EcMot1(Δ CT)–EcTBP density maps. **c**, Schematic of the DNA probes with phosphorothioates (green/grey lollipops) used in FeBABE cleavage assays. **d**, FeBABE-mediated cleavage of Mot1, analysed by western blot³⁰ with approximate sizes of the cleavage products in kilodaltons. **e**, Summary of FeBABE results. Asterisks represent approximate sites of cleavage mediated by FeBABE conjugated to the DNA upstream of the

TATA box. **f**, Model of the Mot1–TBP–DNA complex. Electron density map of EcMot1(BeF)–EcTBP complexes with the crystal structure of EcMot1(NTD)–EcTBP, including a superimposed elongated DNA from the ScTBP–DNA complex (PDB code 1YTB). Bases that represent 5-IdU substitutions used for crosslinking ScMot1 to DNA¹⁵, and bases that represent the FeBABE probe 4Fe (Supplementary Fig. 7a), are coloured in magenta and green, respectively. Positions of FeBABE conjugation that did not produce cleavage are coloured in grey. The position of the Swi2/Snf2 domain of Mot1 is indicated as an orange mesh.

adopts a slightly different conformation, not unexpected for a large HEAT array, a particular lateral density patch was seen to be missing, thereby defining the location of TBP in the complex. Altogether, these data allowed us to rigid-body-dock the Mot1(NTD)-TBP crystal structure convincingly into the electron microscopy density (Supplementary Fig. 6a).

To corroborate this placement, we superimposed TBP in the crystal structure with the ScTBP-DNA complex, and extended the ends of the DNA with generic B-form DNA. Indeed, the upstream DNA protrudes towards the electron density corresponding to the Swi2/Snf2 domain in the electron microscopy three-dimensional reconstruction (Fig. 3f). Our model predicts that the Swi2/Snf2 domain contacts the DNA about 10–17 bases upstream from the TATA sequence, well positioned to translocate along the minor groove of the DNA³. This is in good agreement with previous crosslinking results and satisfactorily explains why a duplex DNA extension is required upstream of the TBP binding site for formation of a catalytically active ScMot1-TBP-DNA complex^{14,15}. To validate this model further, we localized the region of ScMot1 proximal to the upstream DNA using FeBABE-mediated hydroxyl radical cleavage¹⁶ (Fig. 3c and Supplementary Fig. 7a). As predicted by the model, FeBABE molecules positioned within a 9-bp DNA segment immediately upstream of the TATA sequence generated several specific C-terminal Mot1 fragments (cleavage in the Swi2/Snf2 domain), whereas no cleavage products were detected without FeBABE or when FeBABE molecules were conjugated to DNA upstream of this region or downstream of the TATA sequence (Fig. 3d, e and Supplementary Fig. 7b).

Our combined data indicate that Mot1 recognizes TATA-bound TBP by binding to the positively charged TBP surface at H1 and H2, and by binding of the Swi2/Snf2 domain to the minor groove of upstream DNA. We suggest that ATP-dependent groove tracking of the Swi2/Snf2 domain initially disrupts TBP-TATA, followed by binding of the latch to the exposed hydrophobic groove of TBP and full dissociation of Mot1-TBP from DNA (Fig. 4a). In this model, which is consistent with the translocation direction inferred for nucleosome remodelling enzymes¹⁷, the Swi2/Snf2 domain 'pulls' on TBP. Alternatively, the Swi2/Snf2 domain might push TBP. The precise tracking direction must await future studies, although the proposed two-step displacement could occur by translocation in either direction. In any case, the rotational force generated by tracking even a few base pairs of DNA by the Swi2/Snf2 domain could lift TBP from DNA sufficiently for the latch to bind. The energy of a few ATP-dependent translocation steps could be stored elastically in the HEAT repeats. In this way, Mot1 would act like a bottle opener to lift TBP from DNA, with the acidic loops functioning as the head, the HEAT repeats as the handle and the Swi2/Snf2 domain as the twisting hand.

Because TBP exists in many different complexes that could be substrates for the remodelling activity of Mot1, we compared the Mot1-TBP complex with other structurally characterized TBP complexes. The HEAT domain of Mot1 would be able to interact with TBP-TFIIB-DNA complexes as well as with TBP-NC2-DNA complexes (Fig. 4b). The compatibility of Mot1 and NC2 binding to TBP-DNA is consistent with several *in vitro* and *in vivo* results^{8,18–21}, including recent genome-wide chromatin co-localization of Mot1 and NC2 (ref. 20). In contrast, Mot1 sterically overlaps with TFIIA, explaining how Mot1 and TFIIA compete for binding to TBP (Fig. 4c)^{10,11,22}. Mot1 evidently also clashes with Brf1, a subunit of the Pol III initiation factor TFIIB (Fig. 4c), whereas we do not see any clashes with a recent TBP-TFIIB-Pol II preinitiation complex model (Supplementary Fig. 8)^{23,24}. Thus, these comparisons indicate that Mot1 can act on specific subsets of preinitiation complexes in addition to TBP alone. These may include minimal and incomplete preinitiation complexes as well as NC2-repressed TBP complexes, whereas preinitiation complexes that include TFIIA and TBP-associated factors, or Pol III preinitiation complexes (containing Brf1), may be excluded from regulation by Mot1.

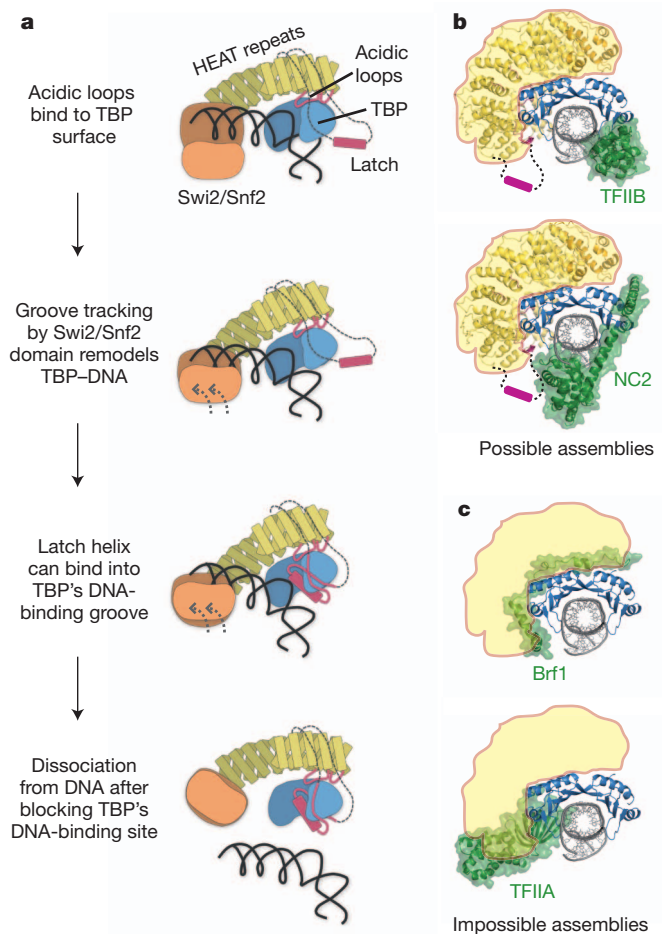


Figure 4 | Proposed remodelling mechanism. **a**, Proposed mechanism of Mot1-mediated displacement of TBP from the DNA. **b**, **c**, Models of possible Mot1 substrates, generated by superimposing the EcMot1-EcTBP crystal structure on other TBP-containing structures. **b**, Possible Mot1 (yellow) substrates are TBP complexes with TFIIB (PDB code 1AIS) and NC2 (PDB code 1JFI). The Mot1 latch is omitted from the structure, but is drawn as a cartoon. **c**, Sterically impossible Mot1 substrates are TBP complexes with TFIIA (PDB code 1NH2) or the TFIIB subunit Brf1 (PDB code 1NGM).

The discovery of the latch and its role in reducing DNA binding and TBP dimerization indicates that Mot1 not only displaces TBP, but blocks its hydrophobic surface patch to prevent interactions with DNA or other factors that bind to the concave surface. Mot1 thus acts as a TBP chaperone to control its interaction with other macromolecules. Mot1 might hold TBP in a diffusible state, explaining how it helps to redistribute TBP rapidly between different promoters and binding sites in the genome. Redistribution between promoters requires large diffusion steps between chromosomes and chromosome loops *in trans*, as opposed to sliding along DNA *in cis*, which is probably part of the repression mode of NC2 (ref. 25). This model is supported by the important role of Mot1 in the high cellular mobility of TBP^{6,7} and by early findings that a substantial proportion of TBP resides in a stable complex with Mot1 in HeLa and yeast cell extracts^{26,27}.

The unusual interactions between Mot1 and TBP might be necessary because of the high-affinity, hydrophobic DNA-binding mode of TBP, as well as the necessity for tight regulation of its binding to specific sites in the genome, while preventing nonspecific DNA interactions. Thus, a combination of motor and chaperone functions could be a more general feature of remodelling systems that deal with the assembly or disassembly of complexes between sticky proteins and DNA. In other systems, remodelling and chaperone functions may be provided by separate factors, as seen, for example, in the cooperation of the SWI/SNF nucleosome remodelling complex and the Asf1 histone chaperone²⁸.

These results provide a high-resolution view of how a Swi2/Snf2-type remodeller interacts with its substrate; they show how the conserved ATP-dependent DNA translocase module can be used to generate high functional specificities within the large and diverse family of Swi2/Snf2 enzymes; and they provide a testable mechanism for a remodelling reaction.

METHODS SUMMARY

Recombinant full-length EcMot1 (residues 1–1275), EcTBP, EcMot1(Δ CT) (residues 1–1016), EcMot1(NTD) (residues 1–779) and EcMot1(Δ latch) (Δ 96–132) were produced in *E. coli* or insect cells. Protein purification was conducted using standard methods and proteins were crystallized by hanging-drop vapour diffusion. EcTBP crystals diffracted to 1.9 Å resolution and were measured at the Swiss Light Source (SLS). Native data of crystals from EcMot1(NTD)–EcTBP showed that they diffracted X-rays to 3.1 Å; these data were collected at the European Synchrotron Radiation Facility (ESRF). Data from derivative crystals of selenomethionine-labelled EcMot1(NTD)–EcTBP were collected to 3.3 Å at the SLS. The structure of EcTBP was solved by molecular replacement using yeast TBP (Protein Data Bank code 1TBP) as a search model. The structure of EcMot1(NTD)–EcTBP was determined using selenium single-wavelength anomalous dispersion in combination with molecular replacement, with the EcTBP structure as a partial model. EcMot1–EcTBP in the presence of 2 mM ADP and beryllium fluoride (ADP-BeF_3^-), EcMot1(E912Q) (the Walker B mutant of EcMot1 was used instead of wild type owing to its enhanced stability) or EcMot1(Δ CT)–EcTBP were used for negative stain (2% uranyl acetate) electron microscopic studies. Micrographs were recorded on a Tecnai G2 Spirit TEM at 120 kV. Size-exclusion experiments were performed on Ettan LC system (GE Healthcare, Superose 12 PC 3.2/30). FeBABE (Dojindo) was conjugated to 68-bp DNA duplexes, based on the sequence of the adenovirus major late promoter. Biotinylation of the top strand's 5' end allowed the duplexes to be bound by streptavidin beads. After FeBABE conjugation, TBP and Mot1 were loaded onto the modified DNAs and cutting was initiated by addition of ascorbic acid and hydrogen peroxide.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 28 January; accepted 18 May 2011.

Published online 6 July 2011.

- Cairns, B. R. The logic of chromatin architecture and remodelling at promoters. *Nature* **461**, 193–198 (2009).
- Li, B., Carey, M. & Workman, J. L. The role of chromatin during transcription. *Cell* **128**, 707–719 (2007).
- Dürr, H., Korner, C., Müller, M., Hickmann, V. & Hopfner, K. P. X-ray structures of the *Sulfolobus solfataricus* SWI2/SNF2 ATPase core and its complex with DNA. *Cell* **121**, 363–373 (2005).
- Saha, A., Wittmeyer, J. & Cairns, B. R. Chromatin remodeling by RSC involves ATP-dependent DNA translocation. *Genes Dev.* **16**, 2120–2134 (2002).
- Racki, L. R. *et al.* The chromatin remodeller ACF acts as a dimeric motor to space nucleosomes. *Nature* **462**, 1016–1021 (2009).
- Auble, D. T. The dynamic personality of TATA-binding protein. *Trends Biochem. Sci.* **34**, 49–52 (2009).
- de Graaf, P. *et al.* Chromatin interaction of TATA-binding protein is dynamically regulated in human cells. *J. Cell Sci.* **123**, 2663–2671 (2010).
- Darst, R. P. *et al.* Mot1 regulates the DNA binding activity of free TATA-binding protein in an ATP-dependent manner. *J. Biol. Chem.* **278**, 13216–13226 (2003).
- Pereira, L. A., van der Knaap, J. A., van den Boom, V., van den Heuvel, F. A. & Timmers, H. T. TAF(II)170 interacts with the concave surface of TATA-binding protein to inhibit its DNA binding activity. *Mol. Cell Biol.* **21**, 7523–7534 (2001).
- Pugh, B. F. Control of gene expression through regulation of the TATA-binding protein. *Gene* **255**, 1–14 (2000).
- Auble, D. T. & Hahn, S. An ATP-dependent inhibitor of TBP binding to DNA. *Genes Dev.* **7**, 844–856 (1993).
- Mohibullah, N. & Hahn, S. Site-specific cross-linking of TBP *in vivo* and *in vitro* reveals a direct functional interaction with the SAGA subunit Spt3. *Genes Dev.* **22**, 2994–3006 (2008).
- Pereira, L. A. *et al.* Molecular architecture of the basal transcription factor B-TFIID. *J. Biol. Chem.* **279**, 21802–21807 (2004).
- Darst, R. P., Wang, D. & Auble, D. T. MOT1-catalyzed TBP–DNA disruption: uncoupling DNA conformational change and role of upstream DNA. *EMBO J.* **20**, 2028–2040 (2001).
- Sprouse, R. O., Brenowitz, M. & Auble, D. T. Snf2/Swi2-related ATPase Mot1 drives displacement of TATA-binding protein by gripping DNA. *EMBO J.* **25**, 1492–1504 (2006).
- Miller, G. & Hahn, S. A DNA-tethered cleavage probe reveals the path for promoter DNA in the yeast preinitiation complex. *Nature Struct. Biol.* **13**, 603–610 (2006).
- Gangaraju, V. K., Prasad, P., Srour, A., Kagalwala, M. N. & Bartholomew, B. Conformational changes associated with template commitment in ATP-dependent chromatin remodeling by ISW2. *Mol. Cell* **35**, 58–69 (2009).
- Dasgupta, A., Darst, R. P., Martin, K. J., Afshari, C. A. & Auble, D. T. Mot1 activates and represses transcription by direct, ATPase-dependent mechanisms. *Proc. Natl Acad. Sci. USA* **99**, 2666–2671 (2002).
- Kleiman, M. P. *et al.* NC2 α interacts with BTA1 and stimulates its ATP-dependent association with TATA-binding protein. *Mol. Cell Biol.* **24**, 10072–10082 (2004).
- van Werven, F. J. *et al.* Cooperative action of NC2 and Mot1p to regulate TATA-binding protein function across the genome. *Genes Dev.* **22**, 2359–2369 (2008).
- Hsu, J. Y. *et al.* TBP, Mot1, and NC2 establish a regulatory circuit that controls DPE-dependent versus TATA-dependent transcription. *Genes Dev.* **22**, 2353–2358 (2008).
- Geisberg, J. V. & Struhl, K. Cellular stress alters the transcriptional properties of promoter-bound Mot1–TBP complexes. *Mol. Cell* **14**, 479–489 (2004).
- Kostrewa, D. *et al.* RNA polymerase II–TFIIB structure and mechanism of transcription initiation. *Nature* **462**, 323–330 (2009).
- Liu, X., Bushnell, D. A., Wang, D., Calero, G. & Kornberg, R. D. Structure of an RNA polymerase II–TFIIB complex and the transcription initiation mechanism. *Science* **327**, 206–209 (2010).
- Schluesche, P., Stelzer, G., Piaia, E., Lamb, D. C. & Meisterernst, M. NC2 mobilizes TBP on core promoter TATA boxes. *Nature Struct. Biol.* **14**, 1196–1201 (2007).
- Timmers, H. T., Meyers, R. E. & Sharp, P. A. Composition of transcription factor B-TFIID. *Proc. Natl Acad. Sci. USA* **89**, 8140–8144 (1992).
- Poon, D., Campbell, A. M., Bai, Y. & Weil, P. A. Yeast Taf170 is encoded by MOT1 and exists in a TATA box-binding protein (TBP)–TBP-associated factor complex distinct from transcription factor IID. *J. Biol. Chem.* **269**, 23135–23140 (1994).
- Glikopoulos, T., Havas, K. M., Dewar, H. & Owen-Hughes, T. SWI/SNF and Asf1p cooperate to displace histones during induction of the *Saccharomyces cerevisiae* HO promoter. *Mol. Cell Biol.* **29**, 4057–4066 (2009).
- Kim, Y., Geiger, J. H., Hahn, S. & Sigler, P. B. Crystal structure of a yeast TBP/TATA-box complex. *Nature* **365**, 512–520 (1993).
- Auble, D. T., Wang, D., Post, K. W. & Hahn, S. Molecular analysis of the SNF2/SWI2 protein family member MOT1, an ATP-driven enzyme that dissociates TATA-binding protein from DNA. *Mol. Cell Biol.* **17**, 4842–4851 (1997).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the Max-Planck Crystallization Facility Martinsried. We thank M. Lucas, A. Schele, C. Ungewickell, J. Goetzl and Y. Hiruma for help with experimentation. We thank J.-P. Armache and M. Turk for help with electron microscopy data. We are grateful to G. Miller and S. Hahn for advice. We thank the staff at the SLS and ESRF for help with data collection. We thank P. Cramer and members of the Hopfner and Auble laboratories for discussions and comments on the manuscript. This work was supported by the German Research Council (SFB 646 and SFB/TR5) and Excellence Initiative (Center for Integrated Protein Science, Munich) to K.-P.H. and R.B., by DFG grant WE4628/1 to P. Wendler and by NIH grant GM55763 to D.T.A.

Author Contributions S.C. and M.M. cloned, purified and crystallized EcTBP; S.C. solved its structure. S.C., A.B., M.M. and P. Wollmann cloned, purified and crystallized EcTBP–EcMot1(NTD); P. Wollmann collected data and P. Wollmann, G.W. and K.-P.H. solved the complex structures. R.V. performed FeBABE experiments, M.N.W. conducted yeast molecular biological manipulations and D.T.A. performed gelshifts. P. Wendler, O.B. and R.B. performed and interpreted electron microscopy experiments. P. Wollmann, P. Wendler, R.B., D.T.A. and K.-P.H. planned and interpreted the experiments. D.T.A. and K.-P.H. wrote the manuscript and all authors provided editorial input.

Author Information Atomic coordinates and structure factors for the reported crystal structures have been deposited with the Protein Data Bank under accession codes 3OC1 (EcTBP) and 3OC3 (EcTBP–EcMot1(NTD) complex). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to K.-P.H. (hopfner@lmb.uni-muenchen.de) or D.T.A. (auble@virginia.edu).

METHODS

Protein preparation. BL21 Rosetta *E. coli* cells (Novagen) were used for expressing EcTBP and EcTBP(K103E) (pET28, Novagen), and for co-expressing EcTBP and EcMot1(NTD) (residues 1–778) or EcMot1(NTD Δ latch) (pET-DUET, Novagen). Proteins were purified by Ni²⁺-affinity chromatography (Qiagen) using a high-salt buffer at pH 8. Further purification of EcTBP was achieved by anion exchange chromatography (HiTrap SP HP, GE Healthcare). For crystallization of EcTBP, the His-tag was removed by tobacco etch virus (TEV) protease digestion. Final purification of all proteins was performed by size exclusion chromatography (Superdex S200, GE Healthcare). Production of selenomethionine-labelled EcTBP and EcMot1(NTD) was done in *E. coli*. Purification of EcTBP and EcMot1(NTD Δ latch), or of selenomethionine-labelled EcTBP and EcMot1(NTD), was performed accordingly.

Sequences encoding full-length EcMot1 (residues 1–1275), EcMot1(Δ CT) (residues 1–1016), EcMot1(NTD) (residues 1–779) and EcMot1(Δ latch) (Δ 96–132), including an N-terminal 10 \times His-tag, were cloned into the pFBDM transfer vector (Invitrogen). EcMot1(E912Q) (Walker B mutant) was generated by site-directed mutagenesis of pFBDM-EcMot1. Transposition of the coding sequence into MultiBac baculoviral DNA was performed in *E. coli* DH10MultiBac^{Cre} cells³¹. Isolated bacmid DNA was used for transfection of *Trichoplusia ni* High Five insect cells (Invitrogen) to produce baculovirus for large-scale infections. Proteins were purified by Ni²⁺-affinity chromatography (Qiagen) using buffer containing 50 mM MES (pH 6.5), 200 mM NaCl, 10 mM β -mercaptoethanol and 12.5 mM or 300 mM imidazole (for EcMot1(NTD) and EcMot1(NTD Δ latch)), or 50 mM Tris (pH 7.5), 400 mM NaCl, 10 mM β -mercaptoethanol, 10% glycerol (v/v) and 12.5 mM or 300 mM imidazole (for EcMot1, EcMot1(Δ CT) and EcMot1(Δ latch)). Additional purification was achieved by ion exchange chromatography (HiTrap Q HP, GE Healthcare). For crystallization of the complex, EcTBP was added in excess amounts to EcMot1. Final purification of the proteins was done by size exclusion chromatography (Superdex S200, GE Healthcare). The preparation of EcMot1(BeF)₃-EcTBP (EcMot1 with ADP-BeF₃[−]) was performed as described previously³².

Crystallization. Proteins were crystallized by hanging-drop vapour diffusion at 18 °C in a mixture of 1 μ l protein (10 mg ml^{−1} EcTBP and 5 mg ml^{−1} EcMot1(NTD)-EcTBP) and 1 μ l precipitant (0.1 M 2-(*N*-morpholino)ethane sulphonic acid (pH 6.5), 2 M NaCl and 4% acetone for EcTBP; 50 mM MES (pH 6), 200 mM ammonium acetate, 5% 2-methyl-2,4-pentanediol, 4% polyethylene glycol 3350 and 200 mM 3-(1-pyridino)-1-propane sulphate (NDSB-201) for EcMot1(NTD)-EcTBP). Crystals were cryoprotected with 1,2-ethanediol (EcTBP) or 2,3-butanediol (EcMot1(NTD)-EcTBP) and flash-frozen in liquid nitrogen.

Structure determination. EcTBP crystals diffracted to 1.9 Å resolution and were measured at the Swiss Light Source (SLS). Native data of crystals from EcMot1(NTD)-EcTBP showed that they diffracted X-rays to 3.1 Å; these data were collected at the European Synchrotron Radiation Facility (ESRF). Data from derivative crystals of selenomethionine-labelled EcMot1(NTD)-EcTBP were collected to 3.3 Å at the SLS. All data were processed with XDS³³. The structure of EcTBP was solved by molecular replacement using Phaser³⁴ and yeast TBP (PDB code 1TBP) as a model. The structure of EcMot1(NTD)-EcTBP was solved by a single anomalous dispersion experiment using SeMet data in combination with molecular replacement, using the EcTBP structure as a partial model (Phaser³⁴). Heavy-atom sites were obtained with SHARP³⁵ and initial automatic model building was performed with Buccaneer³⁶. Model building and refinement was conducted in Coot³⁷ and PHENIX³⁸, respectively. Figures were prepared in Pymol³⁹ or Chimera⁴⁰.

Footprinting assays. Footprinting assays were performed as described previously⁴¹. Reactions contained 20 nM EcTBP, 30 nM EcMot1 and 50 mM ATP, as indicated. After incubation of TBP with DNA for 20 min at 37 °C, Mot1 was added with or without ATP for 5 min before DNase I digestion and sample processing.

ATPase assay. The rates of ATP hydrolysis were measured as described previously⁴¹ in buffer containing 4 mM Tris-HCl (pH 8), 60 mM KCl, 5 mM MgCl₂, 4% glycerol (v/v), 100 mg ml^{−1} BSA and 1 mM dithiothreitol at 22 °C.

Analytical gel filtration. Analytical size exclusion experiments were performed on Ettan LC system (GE Healthcare, Superose 12 PC 3.2/30) according to the manufacturer's instructions (50 mM HEPES (pH 8) or 50 mM MES (pH 6.5), 200 mM NaCl and 2 mM dithiothreitol).

Electrophoretic mobility shift assays. These assays used a radiolabelled fragment of the adenovirus major late promoter⁴¹. Typically, <1 nM DNA was incubated with 15–20 nM EcTBP for 20 min at 37 °C in the same buffer as was used for ATPase assays, then 30 nM Mot1 (or a Mot1 mutant) was added with or without 50 mM ATP for 5 min, before loading on a gel as previously described⁴².

Dissociation kinetic assays. Kinetic analysis of the dissociation reaction was performed by addition of ATP to pre-formed ternary complexes under the

conditions used in the experiment shown in Fig. 2e. Reactions contained radio-labelled DNA (<1 nM), 20 nM EcTBP and either 30 nM wild-type EcMot1 or 30 nM EcMot1(Δ latch). EcTBP was incubated with the radiolabelled DNA template for 20 min, followed by addition of EcMot1 or EcMot1(Δ latch) for 10 min. ATP was added to 100 μ M for 2–20 min and reaction products were resolved at the indicated times on non-denaturing gels. To quantify the extent of complex dissociation at each time point, the free DNA band was quantified and expressed as a proportion of the free DNA present in reactions with no added protein. The results are expressed as the average \pm standard error associated with two independent experiments.

FeBABE cleavage assays. FeBABE (Dojindo) was conjugated to 68-bp DNA duplexes, based on the sequence of the adenovirus major late promoter. Biotinylation of the top strand's 5' end allowed the duplexes to be bound by streptavidin beads. After FeBABE conjugation, TBP and Mot1 were loaded onto the modified DNAs and cutting was initiated by addition of ascorbic acid and hydrogen peroxide. FeBABE-mediated protein cleavage has been previously described^{43,44}. The yeast system was used to take advantage of an antibody raised to the C terminus of yeast Mot1 (ref. 45).

Static light scattering. For molecular weight determination of protein samples (2–4 mg ml^{−1}, with 40 mM HEPES (pH 8), 200 mM NaCl and 2 mM dithiothreitol as a running buffer) by static light scattering, we used a combination of a Viscotek 270 detector and a Viscotek VE-3580 refractive index monitor connected to a microscale HPLC system (AEKTAmicro, GE Healthcare) equipped with an analytical size exclusion column (Superdex S200 15/150 GL, GE Healthcare). Data analysis was performed using the OmniSEC software (Viscotek) using BSA (Thermo Fisher) as a reference for calibration. The chromatographs of the size exclusion, monitored by ultraviolet absorption at 280 nm, and the subsequent refractive index and light-scattering chromatographs all showed a single prominent peak indicating a homogenous sample. Plots of the determined molecular weight versus elution volumes for the evaluated peaks all showed stable molecular weights for the chosen peak areas.

Dynamic light scattering. Dynamic light scattering was measured using a Viscotek/Malvern Instruments 802DLS system. Protein samples (1 mg ml^{−1} in size exclusion buffer) were centrifuged and the supernatant was measured at 20 °C using fluorescence cuvettes. At least ten autocorrelation curves per sample were recorded, averaged and evaluated using the OmniSIZE software and the mass model for globular proteins. All samples showed intensity distributions indicating a homogenous sample with a single peak at the given hydrodynamic radius.

Electron microscopy. 3.5 μ l (10–30 μ g ml^{−1}) of freshly prepared protein sample was applied to pre-coated Quantifoil holey carbon-supported grids and negatively stained using 2% uranyl acetate. Micrographs were recorded on a Tecnai G2 Spirit TEM at 120 kV. Data were collected under low-dose conditions at a nominal magnification of $\times 90,000$ and a nominal defocus of -0.9μ m using an Eagle 2048 \times 2048 pixel CCD camera (FEI Company) with a resolution of 30 μ m pixel^{−1} (3.31 Å pixel^{−1} object scale). 5,518 particles of EcMot1(BeF)₃-EcTBP (EcMot1-EcTBP in the presence of 2 mM ADP and beryllium fluoride), 7,737 of EcMot1(Δ CT)-EcTBP and 12,558 of EcMot1(E912Q) (Walker B mutant of EcMot1 was used instead of wild type owing to its enhanced stability) were picked using boxer⁴⁶. Initial image processing was done using IMAGIC-5 (ref. 47). The images were normalized, filtered at the first zero without CTF correction and centred by iteratively aligning them to their rotationally averaged sum. Initial class averages were obtained by 2–3 rounds of multivariate statistical analysis, followed by multi-reference alignment using homogenous classes as references. The data sets were classified into

10–20 images per class. A low-resolution density map was created by angular reconstruction and was used as an initial model for projection-matching in EMAN 1.9 (ref. 46). The models underwent 8–24 rounds of refinement at an angular increment of up to 5 degrees, until angular assignment was stable. The final reconstructions comprised approximately 90% of the original data set. All visualization and rigid-body fittings were carried out using the UCSF Chimera package⁴⁰. Surface representations show density rendered at a threshold accounting for the expected molecular mass of the complexes: EcMot1(Δ CT)-EcTBP (140 kDa; 170,226 Å³), EcMot1(E912Q) (145 kDa; 172,366 Å³) and EcMot1(BeF)₃-EcTBP (169 kDa; 198,984 Å³). Crystal structures and density maps were merged in VMD using MDFF, as described previously⁴⁸. Difference densities were created by subtracting from one another maps that were rendered at a threshold accounting for the expected molecular mass of the complex.

- Berger, I., Fitzgerald, D. J. & Richmond, T. J. Baculovirus expression system for heterologous multiprotein complexes. *Nature Biotechnol.* **22**, 1583–1587 (2004).
- Kagawa, R., Montgomery, M. G., Braig, K., Leslie, A. G. & Walker, J. E. The structure of bovine F1-ATPase inhibited by ADP and beryllium fluoride. *EMBO J.* **23**, 2734–2744 (2004).
- Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Cryst.* **26**, 795–800 (1993).

34. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674 (2007).
35. Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M. & Paciorek, W. Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Crystallogr. D* **59**, 2023–2030 (2003).
36. Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D* **62**, 1002–1011 (2006).
37. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
38. Adams, P. D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58**, 1948–1954 (2002).
39. DeLano, W. L. *The PyMOL Molecular Graphics System* Version 1.3 r1 (Schrödinger, 2010).
40. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
41. Sprouse, R. O., Brenowitz, M. & Auble, D. T. Snf2/Swi2-related ATPase Mot1 drives displacement of TATA-binding protein by gripping DNA. *EMBO J.* **25**, 1492–1504 (2006).
42. Darst, R. P., Wang, D. & Auble, D. T. MOT1-catalyzed TBP-DNA disruption: uncoupling DNA conformational change and role of upstream DNA. *EMBO J.* **20**, 2028–2040 (2001).
43. Chen, H. T. & Hahn, S. Binding of TFIIB to RNA polymerase II: Mapping the binding site for the TFIIB zinc ribbon domain within the preinitiation complex. *Mol. Cell* **12**, 437–447 (2003).
44. Miller, G. & Hahn, S. A DNA-tethered cleavage probe reveals the path for promoter DNA in the yeast preinitiation complex. *Nature Struct. Mol. Biol.* **13**, 603–610 (2006).
45. Auble, D. T., Wang, D., Post, K. W. & Hahn, S. Molecular analysis of the SNF2/SWI2 protein family member MOT1, an ATP-driven enzyme that dissociates TATA-binding protein from DNA. *Mol. Cell. Biol.* **17**, 4842–4851 (1997).
46. Ludtke, S. J., Baldwin, P. R. & Chiu, W. EMAN: semiautomated software for high-resolution single-particle reconstructions. *J. Struct. Biol.* **128**, 82–97 (1999).
47. van Heel, M., Harauz, G., Orlova, E. V., Schmidt, R. & Schatz, M. A new generation of the IMAGIC image processing system. *J. Struct. Biol.* **116**, 17–24 (1996).
48. Trabuco, L. G., Villa, E., Mitra, K., Frank, J. & Schulten, K. Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. *Structure* **16**, 673–683 (2008).

Excess digestive capacity in predators reflects a life of feast and famine

Jonathan B. Armstrong¹ & Daniel E. Schindler¹

A central challenge for predators is achieving positive energy balance when prey are spatially and temporally heterogeneous. Ecological heterogeneity produces evolutionary trade-offs in the physiological design of predators; this is because the ability to capitalize on pulses of food abundance requires high capacity for food-processing, yet maintaining such capacity imposes energetic costs that are taxing during periods of food scarcity^{1,2}. Recent advances in physiology show that when variation in foraging opportunities is predictable, animals may adjust energetic trade-offs by rapidly modulating their digestive system to track variation in foraging opportunities¹. However, it is increasingly recognized that foraging opportunities for animals are unpredictable³, which should favour animals that maintain a capacity for food-processing that exceeds average levels of consumption (loads)^{2,4}. Despite this basic principle of quantitative evolutionary design, estimates of digestive load:capacity ratios in wild animals are virtually non-existent¹. Here we provide an extensive assessment of load:capacity ratios for the digestive systems of predators in the wild, compiling 639 estimates across 38 species of fish. We found that piscine predators typically maintain the physiological capacity to feed at daily rates 2–3 times higher than what they experience on average. A numerical simulation of the trade-off between food-processing capacity and metabolic cost suggests that the observed level of physiological opportunism is profitable only if predator–prey encounters, and thus predator energy budgets, are far more variable in nature than currently assumed.

Predation opportunities for animals in the wild are distributed heterogeneously in space and time owing to a variety of interactions between predator behaviour⁵ (for example, how they allocate time to foraging, avoiding predators, and finding mates), prey behaviour⁶, and heterogeneity in the physical attributes of habitats where predator–prey interactions occur^{7,8}. As a result, encounter rates between predators and prey in their natural habitats are exceedingly difficult to estimate⁶ and may be far more heterogeneous than assumed. To cope with such ecological heterogeneity, predators employ behavioural and physiological tactics that allow them to store energy when food is plentiful, and utilize such reserves when food is scarce. For predators that store energy internally (for example, as fat reserves or somatic growth), physiological constraints on food-processing may limit rates of energy storage⁹ and compensatory growth¹⁰ and thus the potential to capitalize on pulses of food. To thrive in an environment where foraging opportunities vary widely but unpredictably, predators should maintain physiological opportunism and exhibit maximum capacities for food-processing that considerably exceed the average capacity required.

Acquiring energy from prey is a serial process that spans foraging (attacking, handling, and ingesting prey) to assimilation (digesting food and absorbing nutrients)¹¹. In a homogeneous world, the most economic pathway would exhibit symmorphosis, where the maximal processing rate at each stage is equal to the average input rate (load) from the prior stage¹². However, if the load at any stage varies in time, performance at that stage should exhibit excess capacity, such that the

maximal rate exceeds the mean and the system can accommodate spikes in load⁴. Surprisingly, the role of excess capacity along the food-to-fuel pathway is poorly understood. It is known that parts of the foregut (for example, the oesophagus, stomach, or other comparable structures) can act as food storage reservoirs, enabling predators to ingest prey faster than they can process it^{13,14}. This excess capacity for feeding accommodates short-term variation in prey encounters¹⁴ but it does not facilitate sustained increases in energy gain, which predators would require in order to build energy reserves or achieve compensatory growth. To sustain increased consumption rates, predators require excess capacity for assimilation (that is, digestion and absorption). Current knowledge of excess capacity in the digestive system is primarily limited to measurements at the cellular level², and estimates of load:capacity ratios for integrated rates of assimilation are virtually non-existent for wild animals. Measurements of excess assimilation capacity are important for both physiologists, who seek to evaluate symmorphosis (or lack of it) across the energy intake chain, and ecologists, who lack reliable data to characterize the variation in foraging opportunities experienced by predators in the wild. Here we quantitatively assess how daily variation in foraging opportunity should affect the profitability of excess assimilative capacity, and use our results to interpret observed load:capacity ratios from animals in the wild.

Predators require excess assimilative capacity to capitalize on large pulses of prey abundance as reflected in the positive tail of the distribution of predator foraging opportunities^{8,14}. However, increasing assimilative capacity also increases the energetic costs of physiological maintenance and locomotion¹. We developed a simple numerical simulation to characterize this trade-off and explore its response to the underlying distribution describing daily variation in foraging opportunity. We simulated variation in daily foraging opportunities using a gamma distribution with a fixed arbitrary value for the mean, but different coefficients of variation (c.v., defined as (s.d./mean) × 100) to simulate different levels of variability in the environment (see Methods, Supplementary Fig. 1). A type I functional response¹⁵ modelled how assimilative capacity (C_{\max} , the maximum amount of food that can be consumed in one day) determines the daily food consumption, C , that a predator derives from a daily foraging opportunity, x :

$$C(x) = \begin{cases} x & \text{if } x < C_{\max} \\ C_{\max} & \text{if } x \geq C_{\max} \end{cases} \quad (1)$$

The expected consumption rate ($J d^{-1}$) for an individual with a specified C_{\max} and gamma-distributed variation in daily foraging opportunity is:

$$E(C_{\max}) = \int_0^{\infty} C(x)g(x|k, \theta)dx \quad (2)$$

where $C(x)$ is the type I functional response of equation (1), and $g(x)$ is the gamma distribution with shape and scale parameters (k and θ) varied to produce specified levels of mean and c.v.

¹School of Aquatic and Fishery Sciences, Box 355020, University of Washington, Seattle, Washington 98195, USA.

The load:capacity ratio for assimilation is given by equation (3) below, and its inverse is synonymous with excess capacity, Z :

$$E(C_{\max})/C_{\max} = 1/Z \quad (3)$$

The net energetic profit (N) is the energy remaining after gains from consumption are paid to maintenance and activity costs, M :

$$N = E(C_{\max}) - M \quad (4)$$

The energetic costs associated with increased assimilative capacity are poorly described and are likely to vary among taxa. We model M as a constant fraction, r , of C_{\max} (see Methods):

$$M = rC_{\max}, \text{ where } 0 < r < 1 \quad (5)$$

thus

$$N = E(C_{\max}) - rC_{\max} \quad (6)$$

In our analysis, we used numerical simulation to calculate the load:capacity ratio that maximized the net profit function (equation (6)), given the c.v. in daily foraging opportunity and the cost of the gut. The optimal load:capacity ratio for assimilation decreased with increasing variation in foraging opportunity, but did not exhibit substantial levels of excess capacity until the c.v. was very high, indicating strong right-skew in the distribution of foraging opportunities (Fig. 1 and Supplementary Fig. 1). For example, the optimal load:capacity ratio reached ~ 0.4 when the c.v. in foraging opportunity ranged from 70% to 150%, depending on the costs of maintaining excess capacity (Fig. 1e). This suggests that, in order for integrated rates of assimilation to exhibit levels of excess capacity seen in many other biological structures² (for example, load:capacity ratios < 0.5), predator–prey encounters would need to be extremely heterogeneous in time.

To assess the variation in foraging opportunities that predators experience in nature, we compared our model scenarios to assimilative load:capacity ratios estimated for piscine predators in the wild. We compiled 639 estimates of mean daily consumption rate in fishes. The estimates were calculated from bioenergetics models that integrate across field measurements of growth and express consumption relative to maximum physiological rates. This bioenergetics framework^{16,17} (described in Supplementary Information) has been parameterized for different species through laboratory studies that measure the effects of water temperature and body mass on both metabolism and C_{\max} , the daily consumption rate observed under *ad libitum* feeding. C_{\max} is determined by assimilative capacity¹⁸, and corresponds well to the consumption rates of wild fish in conditions where food is unlimited¹⁹. The model uses field data on predator growth achieved over a defined time period, diet composition and quality, and water temperature to estimate the proportion of C_{\max} (which we call p) that balances an energy budget where growth is surplus energy after energetic gains from consumption are paid to metabolism, excretion and specific dynamic action. Thus, p represents the load:capacity ratio for assimilation. We searched research databases for all papers citing this bioenergetics model and compiled all estimates of p recorded in the wild (Supplementary Table 1). We grouped p estimates by population (unique combinations of species and publication, $n = 66$). The mean length of time over which p was estimated was 208 days (s.d., 143 days).

Across 66 populations from 38 species, the median p was 43% (s.d., 16%), indicating a load:capacity ratio for assimilation of 0.43. The distribution of p was right-skewed and similar in shape whether grouped by population or not (Fig. 2). Populations from marine ($n = 11$), lake ($n = 49$) and stream ($n = 6$) environments did not exhibit significantly different load:capacity ratios (Kruskal–Wallis test: $P = 0.10$, d.f. = 2) and less than 5% of populations exhibited load:capacity ratios greater than 0.8. Among the $>60\%$ of populations that

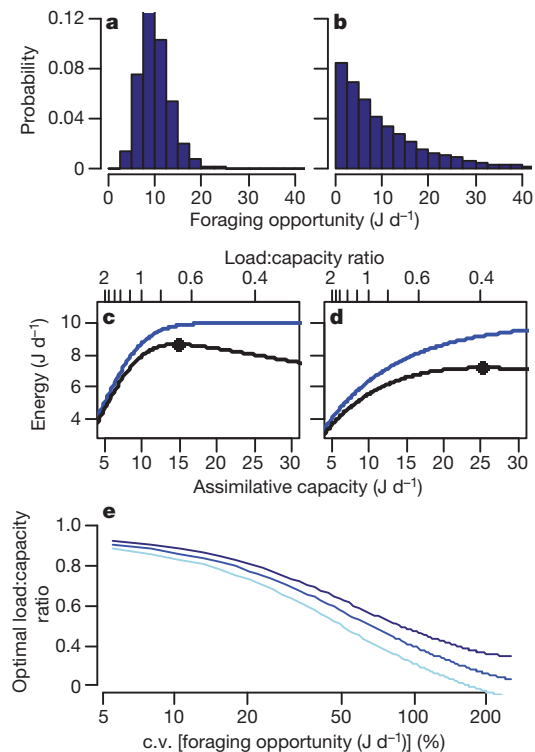


Figure 1 | Results from a simulation model exploring the energetic profitability of excess capacity for assimilation as a function of the daily variation in foraging opportunity. We define assimilation as digestion and absorption. **a, b**, Gamma distributions characterizing two scenarios of ecological heterogeneity that both yield an average of 10 energy units per day in foraging opportunity. **a**, Variance in daily foraging opportunity is equal to the mean, as from a Poisson distribution (c.v. = 32%, $\theta = 1$, $k = 10$). See main text for nomenclature. **b**, Foraging opportunities are highly right-skewed and show 10 times more variance than a Poisson distribution with similar mean (c.v. = 100%, $\theta = 10$, $k = 1$). **c, d**, Cost–benefit analysis of excess assimilative capacity under the two scenarios of ecological heterogeneity; **c** and **d** correspond respectively to distributions displayed in **a** and **b**. Lines represent gross and net energetic gains (blue and black, respectively) resulting from different levels of assimilative capacity (that is, maximum daily consumption rate). The upper x-axis measures excess capacity in terms of the load:capacity ratio (the mean daily consumption rate relative to the maximum rate). Filled circle represents the energetically optimal digestive capacity. Results shown are from the medium cost scenario (see Supplementary Information for a full description). **e**, Model results showing the energetically optimal load:capacity ratio as a function of the c.v. in daily foraging opportunity. Curves depict different cost scenarios (light blue, low; blue, medium; dark blue, high: see Supplementary Information for full description).

exhibited load: capacity ratios less than 0.5, all functional guilds of predators were represented, including planktivores, benthivores and piscivores.

Digestive machinery is expensive^{1,20}, so why do piscine predators maintain the physiological capacity to feed at daily rates that are 2–3 times higher than what they achieve on average? Our numerical simulation demonstrates that such low ratios of assimilative load: capacity become energetically profitable when the distribution of daily foraging opportunities is highly heterogeneous (that is, right-skewed; Fig. 1, Supplementary Fig. 1). This suggests that episodes of gorging and fasting are common in fishes, and occur not only in ambush predators that pursue large prey²¹, but also in predators that feed on insects and zooplankton. These results question common assumptions regarding the variance in foraging opportunities experienced by predators in the wild.

Foraging models frequently simulate variation in a predator's feeding opportunities by drawing prey encounters from the Poisson

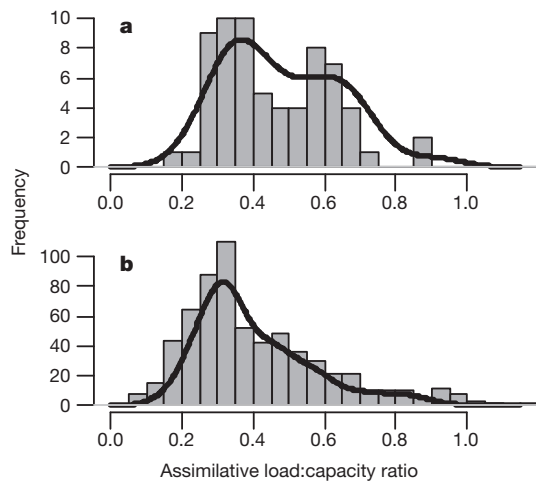


Figure 2 | Histograms showing bioenergetics estimates of integrated consumption rates in wild piscine predators, expressed as the assimilative load:capacity ratio. This ratio is defined as average daily consumption rate divided by the physiological maximum. **a**, Estimates grouped by unique combinations of species and publication ($n = 66$). **b**, All estimates ($n = 639$). Lines show a non-parametric density estimate of each distribution (see Methods).

distribution^{21,22}, which exhibits equal mean and variance (λ). Under most configurations of the Poisson (for example, if $\lambda \geq 3$), values greater than twice the mean have extremely low probabilities (0–3% of occurrences). Thus, foraging models frequently simulate a world in which predators would almost never take advantage of the excess assimilative capacity that we have shown in fishes (that is, median capacity is 2.3 times expected load). This suggests that the feeding opportunities of piscine predators, and perhaps many others, are more heterogeneous than has been appreciated in ecology.

In the past decade, three meta-analyses have concluded that lizards²³, fish²⁴, birds and mammals²⁵ rarely exhibit negative energy budgets at daily timescales, based on observations that predators rarely have completely empty stomachs^{23,24}, or that their integrated rates of energy gain exceed losses at seasonal to annual timescales²⁵. In our analysis, we found that piscine predators probably experience foraging opportunities that are distributed extremely patchily in time, such that the median population would actually spend 18–53% of days feeding at levels that are lower than a typical maintenance ration ($\sim 20\%$ of C_{\max} ; ref. 17). This suggests that fishes routinely run negative energy budgets and warns that prior syntheses of empirical data may have underestimated variability in the daily energy balance of not only fishes, but other taxa as well. Future work should consider how patterns of capacity through the food–fuel pathway are associated with different strategies of energy regulation. For example, many small birds and mammals hoard food during resource pulses or employ torpor to avoid energy deficits²⁶. These mechanisms may stabilize consumption rates and energy budgets enough to make maintaining excess capacity for assimilating food unnecessary, but would probably require increased capacity for handling food.

We have assumed that the assimilative capacity observed in fishes has evolved to optimize energy budgets under trade-offs between the capacity and metabolic cost of the gut. There is widespread evidence that the energetic profitability of the gut has strong fitness consequences^{1,20} and that its morphology and function evolve in response to energetic demands²⁷. Further, the fishes in our analyses exhibit indeterminate somatic growth that is indeed limited by energy acquisition²⁸. Therefore it is reasonable to use an energy maximization model to interpret the function of physiological capacity. The temporal pattern of foraging opportunities may vary among generations such that the optimal level of digestive capacity is not static in time. Our general conclusions are robust to such evolutionary disequilibrium because we found high levels

of excess digestive capacity across numerous taxa inhabiting diverse environments. The most parsimonious explanation of these findings is that piscine predators have evolved excess assimilative capacity to profit in a world where binging and fasting are regular occurrences.

Phenotypic flexibility can improve performance if energy budgets vary predictably¹. For example, female mammals increase digestive capacity during pregnancy, and birds adjust digestive capacity during migration as they alternate between fuelling and flight¹. Although seasonal changes in food abundance may be predictable, pulses of food are frequently both unpredictable and ephemeral²⁹, causing the performance of consumers to be measured by the guts they possess in the moment and not the ones they can construct days or weeks later (except certain sit-and-wait predators that are capable of rapid intestinal regulation²⁰). Similarly to the fishes in our analysis, birds may maintain excess capacity for consumption despite their ability to modify gut size over relatively short time-scales³⁰. Flexibility in gut size has been suggested³⁰ to facilitate adaptive responses to long-term but not short-term variation in resource abundance. Our meta-analysis and model strongly suggest that predator–prey encounters are patchier than most ecologists assume, that fish run negative energy budgets more frequently than assumed, and that eco-physiology should consider the combined roles of excess capacity and phenotypic flexibility when considering how organisms cope with a world far more heterogeneous than the laboratory or treadmill. The insights derived from our analysis have important implications for models of predator–prey dynamics, community structure and the stability of food webs, which can be sensitive to heterogeneity in predator–prey interactions⁶.

METHODS SUMMARY

To characterize the assimilative load:capacity ratios of piscine predators, we compiled data from bioenergetics analyses that express integrated consumption rates relative to the physiological maximum. These models^{16,17} use field measurements of growth at monthly to annual timescales to reconstruct consumption rates, based on the energy budget:

$$G = C - (R + W + \text{SDA}) \quad (7)$$

where G is growth, C is food consumption, R is active metabolism, W is waste (faeces and urine) and SDA is specific dynamic action (the cost of digestion). Empirically derived, species-specific functions model the effects of water temperature and body mass on respiration and the maximum daily consumption rate (C_{\max}). Additional functions model waste and SDA as taxes on consumption. To incorporate thermal and allometric constraints on C , the parameter p scales the actual consumption rate relative to the maximum consumption rate according to $p = C/C_{\max}$. Thus, equation (7) becomes:

$$G = pC_{\max} - (R + W + \text{SDA}) \quad (8)$$

On the basis of observed growth, diet composition, predator and prey energy densities, and the temperature regime experienced by a fish for the time interval over which growth is estimated, the model solves for the value of p that produces the value of G observed in the field. Because C_{\max} is determined by rates of food-processing¹⁸, p represents the load:capacity ratio for assimilation—the process that includes both the break down of food macromolecules and the transportation of nutrients across the gut wall.

For our meta-analysis of p estimates, we searched Web of Science and Google Scholar for all publications ($n = 345$) citing the bioenergetics framework outlined above^{16,17}. We compiled data from all publications (42 journal articles and 4 theses) that studied predators *in situ* and reported the parameter p . The metadata and source publications are provided in Supplementary Table 1.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 7 March; accepted 27 May 2011.

Published online 6 July 2011.

1. Piersma, T. & van Gils, J. A. *The Flexible Phenotype* (Oxford Univ. Press, 2011).
2. Diamond, J. Quantitative evolutionary design. *J. Physiol. (Lond.)* **542**, 337–345 (2002).

3. Humphries, N. E. *et al.* Environmental context explains Levy and Brownian movement patterns of marine predators. *Nature* **465**, 1066–1069 (2010).
4. Gans, C. Momentarily excessive construction as the basis for protoadaptation. *Evolution* **33**, 227–233 (1979).
5. Ritchie, E. G. & Johnson, C. N. Predator interactions, mesopredator release and biodiversity conservation. *Ecol. Lett.* **12**, 982–998 (2009).
6. Walters, C. J. M. & Martell, S. J. D. *Fisheries Ecology and Management* (Princeton Univ. Press, 2004).
7. Kauffman, M. J. *et al.* Landscape heterogeneity shapes predation in a newly restored predator-prey system. *Ecol. Lett.* **10**, 690–700 (2007).
8. Armstrong, J. B. *et al.* Thermal heterogeneity mediates the effects of pulsed subsidies across a landscape. *Ecology* **91**, 1445–1454 (2010).
9. Klaassen, M., Lindstrom, A. & Zijlstra, R. Composition of fuel stores and digestive limitations to fuel deposition rate in the long-distance migratory thrush nightingale, *Luscinia luscinia*. *Physiol. Zool.* **70**, 125–133 (1997).
10. Dupont-Prinet, A. *et al.* Physiological mechanisms underlying a trade-off between growth rate and tolerance of feed deprivation in the European sea bass (*Dicentrarchus labrax*). *J. Exp. Biol.* **213**, 1143–1152 (2010).
11. Weiner, J. Physiological limits to sustainable energy budgets in birds and mammals: ecological implications. *Trends Ecol. Evol.* **7**, 384–388 (1992).
12. Weibel, E. R., Taylor, C. R. & Hoppeler, H. The concept of symmorphosis: a testable hypothesis of structure-function relationship. *Proc. Natl Acad. Sci. USA* **88**, 10357–10361 (1991).
13. Kersten, M. & Visser, W. The rate of food processing in the oystercatcher: food intake and energy expenditure constrained by a digestive bottleneck. *Funct. Ecol.* **10**, 440–448 (1996).
14. Essington, T. E., Hodgson, J. R. & Kitchell, J. F. Role of satiation in the functional response of a piscivore, largemouth bass (*Micropterus salmoides*). *Can. J. Fish. Aquat. Sci.* **57**, 548–556 (2000).
15. Holling, C. S. The components of predation as revealed by a study of small-mammal predation of the European pine sawfly. *Can. Entomol.* **91**, 293–320 (1959).
16. Kitchell, J. F., Stewart, D. J. & Weininger, D. Applications of a bioenergetics model to yellow perch (*Perca flavescens*) and walleye (*Stizostedion vitreum-vitreum*). *J. Fish. Res. Board Can.* **34**, 1922–1935 (1977).
17. Fish Bioenergetics 3.0 (University of Wisconsin-Madison Centre for Limnology/Wisconsin Sea Grant Institute, 1997).
18. Elliott, J. M. & Persson, L. Estimation of daily rates of food consumption for fish. *J. Anim. Ecol.* **47**, 977–991 (1978).
19. Kitchell, J. F. *et al.* Predator-prey dynamics in an ecosystem context. *J. Fish Biol.* **45**, 209–226 (1994).
20. Secor, S. M., Stein, E. D. & Diamond, J. Rapid up-regulation of snake intestine in response to feeding: a new model of intestinal adaptation. *Am. J. Physiol.* **266**, G695–G705 (1994).
21. Breck, J. E. Foraging theory and piscivorous fish: are forage fish just big zooplankton? *Trans. Am. Fish. Soc.* **122**, 902–911 (1993).
22. DeAngelis, D. L. & Gross, L. J. *Individual-based Models and Approaches in Ecology. Populations, Communities, and Ecosystems* (Chapman and Hall, 1992).
23. Huey, R. B., Pianka, E. R. & Vitt, L. J. How often do lizards “run on empty”? *Ecology* **82**, 1–7 (2001).
24. Arrington, D. A., Winemiller, K. O., Loftus, W. F. & Akin, S. How often do fishes “run on empty”? *Ecology* **83**, 2145–2151 (2002).
25. Jeschke, J. M. When carnivores are “full and lazy”. *Oecologia* **152**, 357–364 (2007).
26. Brodin, A. & Clark, C. in *Foraging* (eds Stephen, D. S., Brown, J. S. & Ydenberg, R. C.) 221–269 (Univ. Chicago Press, 2007).
27. Hofmann, R. R. Evolutionary steps of ecophysiological adaptation and diversification of ruminants: a comparative view of their digestive system. *Oecologia* **78**, 443–457 (1989).
28. Schindler, D. E. & Eby, L. A. Stoichiometry of fishes and their prey: implications for nutrient recycling. *Ecology* **78**, 1816–1831 (1997).
29. Yang, L. H., Bastow, J. L., Spence, K. O. & Wright, A. N. What can we learn from resource pulses? *Ecology* **89**, 621–634 (2008).
30. McWilliams, S. R. & Karasov, W. H. Phenotypic flexibility in digestive system structure and function in migratory birds and its ecological significance. *Comp. Biochem. Physiol. A* **128**, 577–593 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank T. Essington, R. Huey, T. Reed, A. Walters, V. Sturtevant and A. Armstrong for comments on this manuscript. We also thank the following people for contributing to this project: J. Kitchell, O. Jensen, E. Ward, D. Beauchamp, B. Chasco, A. Farrell, P. Bisson and J. Kershner. This work was supported by the Gordon and Betty Moore Foundation, the US National Science Foundation and the University of Washington School of Aquatic and Fishery Sciences.

Author Contributions J.B.A. and D.E.S. contributed to each stage of the project.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.B.A. (Jonny99@uw.edu).

METHODS

The fish bioenergetics model (FBEM). This model^{16,17} is used to estimate assimilative load:capacity ratios. It uses an energy balance approach to bioenergetics that provides an adaptable modelling framework, which has been applied to a wide variety of species and ecosystems. The model balances an energy budget where inputs must equal outputs, specifically:

$$C = G + R + W + \text{SDA} \quad (9)$$

Here C is food consumption, G is growth, R is active metabolism, W is waste (faeces and urine), and SDA is specific dynamic action (the cost of digestion). The FBEM provides functions that model the effects of water temperature and body mass on respiration and the maximum daily consumption rate (C_{\max}). These functions are parameterized for individual species through controlled laboratory experiments. To formulate the respiration functions, oxygen consumption is measured in experiments that manipulate body size and water temperature. In addition, an activity multiplier is calculated to relate basal and active metabolism. To formulate the C_{\max} functions, daily consumption rates are measured in *ad libitum* feeding experiments that manipulate temperature and body size. Additional functions model waste and SDA as taxes on consumption. A full description of these functions is available in the Bioenergetics 3.0 software manual¹⁷.

Researchers typically use the FBEM to estimate consumption rates based on the observed growth rates of fishes in the field. This is done by rearranging the original terms in the energy budget so that growth represents surplus energy after gains from consumption are paid to energetic costs:

$$G = C - (R + W + \text{SDA}) \quad (10)$$

Growth is measured directly over some relatively long time interval (for example, monthly to annual growth increments) and the model is used to estimate the consumption rate C required to satisfy equation (10) to produce the observed growth increment. To incorporate thermal and allometric constraints on C , the parameter p scales the actual consumption rate relative to the maximum consumption rate according to $p = C/C_{\max}$. Thus, equation (10) becomes:

$$G = pC_{\max} - (R + W + \text{SDA}) \quad (11)$$

On the basis of observed growth, diet composition, predator and prey energy densities, and the temperature regime experienced by a fish for the time interval over which growth is estimated, the model solves for the value of p that produces the value of G observed in the field. The model is implemented at a daily time step and linearly interpolates between observations of temperature, diet composition and energy density. Because growth of fishes is indeterminate, the value of p is a sensitive but informative integrated parameter that reflects the realized consumption rate relative to the physiological maximum rate that an individual fish is capable of, based on its metabolic parameters, its body size, its energy density and the energy density of its prey, and water temperature. Because C_{\max} is determined by rates of food-processing¹⁸, p represents the load:capacity ratio for assimilation—the process that includes both the break-down of food macromolecules and the transportation of nutrients across the gut wall. Consumption estimates from the FBEM have been independently corroborated by intensive field-based methods^{31,32}.

For our meta-analysis of p estimates, we searched Web of Science and Google Scholar for all publications ($n = 345$) citing the bioenergetics framework outlined above^{16,17}. We compiled data from all publications that studied predators *in situ* and reported the parameter p (42 journal articles and 4 theses). The meta-data and source publications are provided in Supplementary Table 1. To characterize distributions of p estimates (Fig. 2), we fitted kernel density estimates to the data in R³³. We used a Gaussian smoothing kernel and selected the bandwidth following standard methods³⁴.

Modelling trade-offs associated with digestive capacity. Here we give details of our numerical model in which we simulate these trade-offs. We used the gamma distribution to simulate variation in daily foraging opportunity because it can generate a diversity of biologically realistic shapes, ranging from narrow and

Gaussian to extremely right-skewed, representing an environment where prey encounters are extremely patchy. The gamma distribution has two parameters, namely k and θ , which relate to the mean and variance of the distribution as follows:

$$\begin{aligned} \bar{x} &= k\theta \\ \sigma^2 &= k\theta^2 \end{aligned} \quad (12)$$

Rearranging equation (12) yields:

$$\begin{aligned} k &= \bar{x}^2/\sigma^2 \\ \theta &= \sigma^2/\bar{x} \end{aligned} \quad (13)$$

In order to examine the effect of variability in foraging opportunity on the energetic profitability of excess assimilative capacity, we varied θ and k to generate gamma distributions with an equal mean, but different levels of variation. This simulates ecosystems that have an equal amount of prey available to predators, but different levels of variation in the temporal patterning of daily predator-prey encounters.

To provide biologically interpretable results, we used the coefficient of variation (c.v., defined as (s.d./mean) \times 100) to describe variation in the gamma distribution. Unlike other metrics of variation (for example, the variance to mean ratio) the relationship between the c.v. and the shape of the gamma distribution scales isometrically with the mean of the distribution (for example, if the c.v. is held constant, changing the mean does not affect the skew of the distribution). Therefore, we could explore all ecologically relevant shapes of the gamma distribution by changing the c.v. of the distribution (range, 0.1–500%) while keeping the mean fixed at an arbitrary value. Supplementary Fig. 1 shows a subset of the gamma distributions that were generated in our simulations to illustrate the diversity of shapes that we considered.

The relationship between the maintenance cost and assimilative capacity of the gut (R_{gut} and C_{\max}) is poorly documented. For the sake of parsimony, and because empirical data suggest that assimilative capacity in fish is related to the surface area of the gut³⁵, we modelled a linear relationship between C_{\max} and R_{gut} , such that a proportional change in C_{\max} produces the same proportional change in R_{gut} (for example, doubling C_{\max} would double R_{gut}). The relationship between C_{\max} and total metabolic cost, M , depends on the ratios of $R_{\text{gut}}:M$ and $M:C_{\max}$. Under the assumed linear relationship, the slope (r) of M as a function of C_{\max} is:

$$r = R_{\text{gut}}/R_{\text{tot}} \times R_{\text{tot}}/C_{\max} \quad (14)$$

We assumed that one-third of total daily respiration goes to maintaining digestive capacity, as this is the approximate proportion of total cardiac output delivered to an empty gut³⁶. The empirically derived functions in the FBEM¹⁷ suggest energetic losses due to respiration typically represent one-third to one-ninth of C_{\max} in fishes. Given these estimations, r ranges from 0.04 to 0.11. Because the y -intercept of a linear cost function (here, the cost of metabolism not associated with the gut) does not affect the optimal value for profit maximization, we modelled total metabolic cost, M , as a constant fraction, r , of C_{\max} , and included three cost scenarios: low ($r = 4\%$), medium ($r = 8\%$) and high ($r = 12\%$). We did not consider specific dynamic action or excretion in our cost analysis, because we assumed these were taxes on consumption that do not represent physiological maintenance.

31. Rice, J. A. & Cochran, P. A. Independent evaluation of a bioenergetics model for largemouth bass. *Ecology* **65**, 732–739 (1984).
32. Beauchamp, D. A., Stewart, D. J. & Thomas, G. L. corroboration of a bioenergetics model for sockeye salmon. *Trans. Am. Fish. Soc.* **118**, 597–607 (1989).
33. R Development Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna); available at (<http://www.R-project.org>) (2010).
34. Scott, D. W. *Multivariate Density Estimation: Theory, Practice, and Visualization* (Wiley, 1992).
35. Stevens, E. D. & Devlin, R. H. Intestinal morphology in growth hormone transgenic coho salmon. *J. Fish Biol.* **56**, 191–195 (2000).
36. Farrell, A. P. *et al.* Gut blood flow in fish during exercise and severe hypercapnia. *Comp. Biochem. Physiol. A* **128**, 549–561 (2001).

Sideband cooling of micromechanical motion to the quantum ground state

J. D. Teufel¹, T. Donner^{2,3}, Dale Li¹, J. W. Harlow^{2,3}, M. S. Allman^{1,3}, K. Cicak¹, A. J. Sirois^{1,3}, J. D. Whittaker^{1,3}, K. W. Lehnert^{2,3} & R. W. Simmonds¹

The advent of laser cooling techniques revolutionized the study of many atomic-scale systems, fuelling progress towards quantum computing with trapped ions¹ and generating new states of matter with Bose–Einstein condensates². Analogous cooling techniques^{3,4} can provide a general and flexible method of preparing macroscopic objects in their motional ground state. Cavity optomechanical or electromechanical systems achieve sideband cooling through the strong interaction between light and motion^{5–15}. However, entering the quantum regime—in which a system has less than a single quantum of motion—has been difficult because sideband cooling has not sufficiently overwhelmed the coupling of low-frequency mechanical systems to their hot environments. Here we demonstrate sideband cooling of an approximately 10-MHz micromechanical oscillator to the quantum ground state. This achievement required a large electromechanical interaction, which was obtained by embedding a micromechanical membrane into a superconducting microwave resonant circuit. To verify the cooling of the membrane motion to a phonon occupation of 0.34 ± 0.05 phonons, we perform a near-Heisenberg-limited position measurement³ within $(5.1 \pm 0.4)\hbar/2\pi$, where \hbar is Planck's constant. Furthermore, our device exhibits strong coupling, allowing coherent exchange of microwave photons and mechanical phonons¹⁶. Simultaneously achieving strong coupling, ground state preparation and efficient measurement sets the stage for rapid advances in the control and detection of non-classical states of motion^{17,18}, possibly even testing quantum theory itself in the unexplored region of larger size and mass¹⁹. Because mechanical oscillators can couple to light of any frequency, they could also serve as a unique intermediary for transferring quantum information between microwave and optical domains²⁰.

A mechanical oscillator of high quality factor placed within the quantum regime could allow us to explore quantum mechanics in entirely new ways^{17–20}. To do this requires the ability to prepare the oscillator in its ground state, to arbitrarily control its quantum states, and to detect these states near the Heisenberg limit. In addition, the oscillator system should not be strongly perturbed by its environment or any other extraneous influence, including dissipation or thermal excitations. As a first step, the oscillator's temperature T must be reduced so that $k_B T < \hbar\Omega_m$, where Ω_m is the resonance frequency of the oscillator, k_B is Boltzmann's constant, and \hbar is $\hbar/2\pi$. Although there has been substantial progress in cooling mechanical oscillators with radiation pressure forces, sideband cooling to the quantum mechanical ground state has been a long-standing challenge. Cavity optomechanical systems have realized very large sideband cooling rates^{9–13,15}; however, these rates are not sufficient to overcome the larger thermal heating rates of the mechanical modes. Electromechanical experiments using much lower-energy microwave photons^{5–8,14}, although simpler to operate below 100 mK, have suffered from weak electromechanical interactions and inefficient detection of the photon fields.

In a unique approach, a system based on a high-frequency (6-GHz) microwave dilatation oscillator was integrated with a superconducting

phase qubit²¹. Its high frequency offered the advantage of reaching the ground state at relatively high temperatures ($T \approx 25$ mK), which were achievable simply with passive dilution refrigeration. Furthermore, because the mechanical oscillator was piezoelectric, its strong electrical response enabled significantly strong coupling to the superconducting qubit, providing direct control and (destructive) measurement of the phonon energy states. These results showed that quantum effects are achievable with a human-fabricated mechanical oscillator. Unfortunately, the short mechanical lifetimes prevented the manipulation of complex mechanical states and direct tests of entanglement.

Low frequency (<100 -MHz) mechanical oscillators have distinct advantages: higher quality factors, long phonon lifetimes and large motional state displacements, which are important for future tests of quantum theory¹⁹. Cavity opto- or electro-mechanical systems⁴ naturally offer a powerful method for both cooling and detecting low-frequency mechanical oscillators^{22,23}. An object whose motion alters the resonance frequency, ω_c , of an electromagnetic cavity experiences a radiation pressure force governed by the parametric interaction Hamiltonian: $\hat{H}_{\text{int}} = \hbar G \hat{n} \hat{x}$, where $G = d\omega_c/dx$, \hat{n} is the cavity photon number, and \hat{x} is the displacement of the mechanical oscillator. By driving the cavity at a frequency ω_d , the oscillator's motion produces upper and lower sidebands at $\omega_d \pm \Omega_m$. Because these sideband photons are inelastically scattered from the drive field, they provide a way to exchange energy with the oscillator. If the drive field is optimally detuned below the cavity resonance by an amount $\Delta \equiv \omega_d - \omega_c = -\Omega_m$, photons will be preferentially up-converted to ω_c because the photon density of states is maximal there (Fig. 1b). When an up-converted photon leaves the cavity, it removes the energy of one mechanical quantum (one phonon) from the motion. Thus, the mechanical oscillator is damped and cooled by way of this radiation-pressure force. Because the mechanical motion is encoded in the scattered photons leaving the cavity, information on the position of the mechanical oscillator provides a near Heisenberg-limited measurement of displacement²⁴.

Here we use a cavity electromechanical system where a flexural mode of a thin aluminium membrane is parametrically coupled to a superconducting microwave resonant circuit. Unlike previous microwave systems, this device achieves large electromechanical coupling by use of a flexible vacuum-gap capacitor^{16,25}. The oscillator is a 100-nm-thick aluminium membrane with a diameter of 15 μm , suspended 50 nm above a second aluminium layer on a sapphire substrate²⁵ (Fig. 1). These two metal layers form a parallel-plate capacitor that is shunted by a 12-nH spiral inductor. This combination of capacitor and inductor creates a microwave cavity with a displacement-dependent resonance frequency centred at $\omega_c = 2\pi \times 7.54$ GHz. The device is operated in a dilution refrigerator at 15 mK, at which temperature aluminium is superconducting, and the microwave cavity has a total energy decay rate of $\kappa \approx 2\pi \times 200$ kHz. The diameter of the aluminium membrane and its tension⁸ produce an Ω_m of $2\pi \times 10.56$ MHz with an intrinsic damping rate of $\Gamma_m = 2\pi \times 32$ Hz, or mechanical quality factor $Q_m = \Omega_m/\Gamma_m = 3.3 \times 10^5$. The oscillator mass, $m = 48$ pg, implies that

¹National Institute of Standards and Technology (NIST), Boulder, Colorado 80305, USA. ²JILA, University of Colorado and NIST, Boulder, Colorado 80309, USA. ³Department of Physics, University of Colorado, Boulder, Colorado 80309, USA.

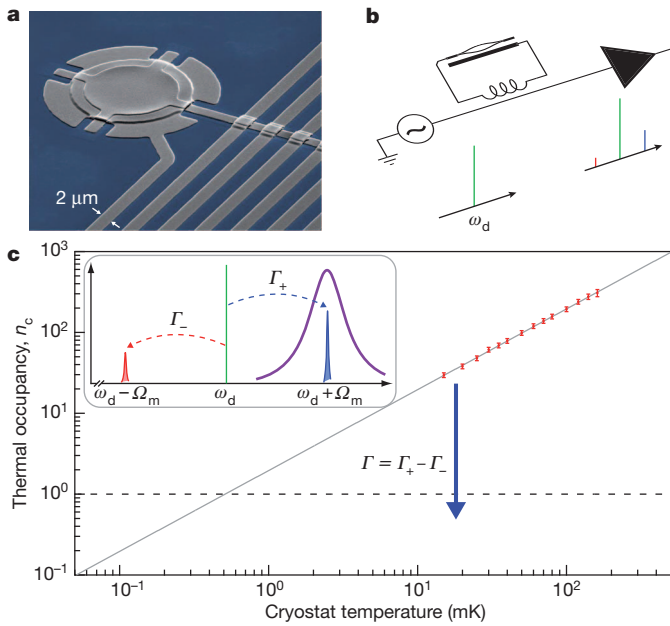


Figure 1 | Schematic description of the experiment. **a**, False-colour scanning electron micrograph showing the aluminium (grey) electromechanical circuit fabricated on a sapphire (blue) substrate; a 15- μm -diameter membrane is suspended 50 nm above a lower electrode. The membrane's motion modulates the capacitance, and hence, the resonance frequency of the superconducting microwave circuit. **b**, A coherent microwave drive (left, ω_d , shown green in frequency-amplitude plot below) inductively coupled to the circuit (top) acquires modulation sidebands (red and blue in plot below) owing to the thermal motion of the membrane. The upper sideband is amplified with a nearly quantum-limited Josephson parametric amplifier (filled triangle, right) within the cryostat. **c**, The microwave power in the upper sideband provides a direct measurement of the thermal occupancy of the mechanical mode, which may be calibrated *in situ* by varying the temperature of the cryostat (main panel). The mechanical mode shows thermalization with the cryostat at all temperatures, yielding a minimum thermal occupancy of 30 mechanical quanta without using sideband-cooling techniques. Error bars, s.d. Inset, illustration of the concept of sideband cooling. When the circuit is excited with a detuned microwave drive such that $\Delta = -\Omega_m$, the narrow line shape of the electrical resonance ensures that the rate to scatter photons to higher energy Γ_+ (blue dashed arrow, blue peak) exceeds the rate to scatter to lower energy Γ_- (red dashed arrow, red peak). Thus, the net scattering rate Γ (blue solid arrow) provides a cooling mechanism for the membrane.

the zero-point motion is $x_{zp} = \sqrt{\hbar/(2m\Omega_m)} = 4.1$ fm. With a ratio of $\Omega_m/\kappa > 50$, our system is deep within the resolved-sideband regime and well-suited for sideband cooling to the mechanical ground state^{22,23}.

To measure the mechanical displacement, we apply a microwave field, which is detuned below the cavity resonance frequency by $\Delta = -\Omega_m$, through heavily attenuated coaxial lines to the feed line of our device. The upper sideband, now at ω_c , is amplified with a custom-built Josephson parametric amplifier^{26,27} followed by a low-noise cryogenic amplifier, demodulated at room temperature, and finally monitored with a spectrum analyser. The thermal motion of the membrane creates an easily resolvable peak in the microwave noise spectrum. As described previously²⁷, this measurement scheme constitutes a nearly shot-noise-limited microwave interferometer with which we can measure mechanical displacement with minimum added noise close to fundamental limits.

In order to calibrate the demodulated signal to the membrane's motion, we measure the thermal noise spectrum while varying the cryostat temperature (Fig. 1c). Here a weak microwave drive (~ 3 photons in the cavity) is used in order to ensure that radiation pressure damping and cooling effects are negligible. When $\Omega_m \gg \kappa \gg \Gamma_m$ and $\Delta = -\Omega_m$, the displacement spectral density S_x is related to the observed microwave noise spectral density S by $S_x = 2(\kappa\Omega_m/G\kappa_{\text{ex}})^2 S/P_o$, where κ_{ex} is the

coupling rate between the cavity and the feed line, and P_o is the power of the microwave drive at the output of the cavity. According to equipartition, the area under the resonance curve of displacement spectral density S_x must be proportional to the effective temperature of the mechanical mode. This calibration procedure allows us to convert the sideband in the microwave power spectral density to a displacement spectral density and to extract the thermal occupation of the mechanical mode. In Fig. 1c we show the number of thermal quanta in the mechanical resonator as a function of T . The linear dependence of the integrated power spectral density with temperature shows that the mechanical mode equilibrates with the cryostat even for the lowest achievable temperature of 15 mK. This temperature corresponds to a thermal occupancy $n_m = 30$, where $n_m = [\exp(\hbar\Omega_m/k_B T) - 1]^{-1}$. The calibration determines the electromechanical coupling strength, $G/2\pi = 49 \pm 2$ MHz nm⁻¹. With these device parameters, we can investigate both the fundamental sensitivity of our measurement and the effects of radiation pressure cooling.

The total measured displacement noise results from two sources: the membrane's actual mean-square motion, S_x^{th} , and its apparent motion, S_x^{imp} , which is due to imprecision of the measurement. Figure 2a demonstrates how the use of low-noise parametric amplification significantly

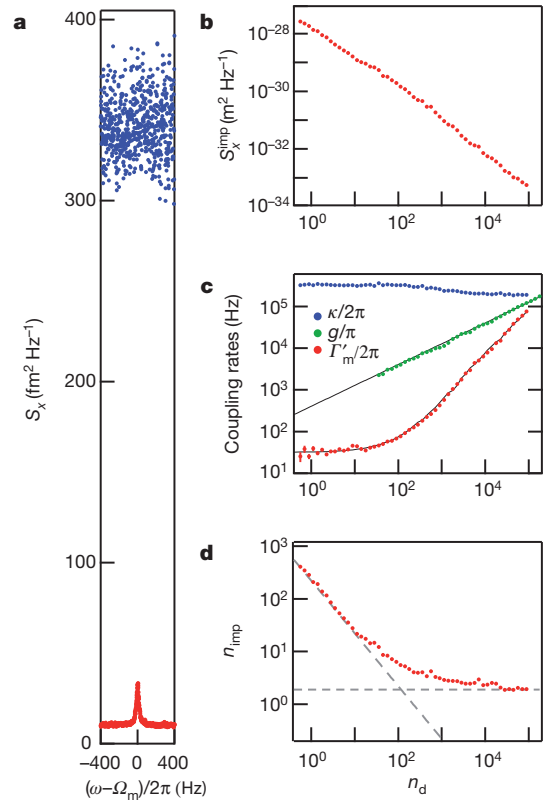


Figure 2 | Displacement sensitivity in the presence of dynamical back-action. **a**, The displacement spectral density S_x measured with (red) and without (blue) the Josephson parametric amplifier. As the parametric amplifier greatly reduces the total noise of the microwave measurement, the time required to resolve the thermal motion is reduced by a factor of 1,000. **b**, As the microwave drive power is increased, the absolute displacement sensitivity, S_x^{imp} improves, reaching a minimum of 5.5×10^{-34} m² Hz⁻¹ at the highest power. **c**, The parametric coupling rate g between the microwave cavity and the mechanical mode increases as $\sqrt{n_d}$. This coupling broadens the linewidth of the mechanical mode Γ'_m from its intrinsic value of $\Gamma_m = 2\pi \times 32$ Hz until it exceeds the linewidth of the microwave cavity κ . **d**, The relative measurement imprecision, in units of mechanical quanta, depends on the product of S_x^{imp} and Γ'_m . Thus, once the power is large enough that dynamical back-action overwhelms the intrinsic mechanical linewidth, n_{imp} asymptotically approaches a constant value ($n_{\text{imp}} = 1.9$), which is a direct measure of the overall efficiency of the photon measurement.

lowers S_x^{imp} , resulting in a reduction in the white-noise background by a factor of more than 30. This greatly increases the signal-to-noise ratio of the membrane's thermal motion, thereby reducing the integration time required to resolve the thermal peak by a factor of 1,000. To investigate the measurement sensitivity in the presence of dynamical back-action, we regulate the cryostat temperature at 20 mK and increase the amplitude of the detuned microwave drive while observing modifications in the displacement spectral density. We quantify the strength of the drive by the resulting number of photons n_d in the microwave cavity. As shown in Fig. 2b, the measurement imprecision S_x^{imp} is inversely proportional to n_d . At the highest drive power ($n_d \approx 10^5$), the absolute displacement sensitivity is $5.5 \times 10^{-34} \text{ m}^2 \text{ Hz}^{-1}$.

As expected, the increased drive power also damps and cools the mechanical oscillator^{3,22,23}. The total mechanical dissipation rate $\Gamma'_m = \Gamma_m + \Gamma$ is the sum of the intrinsic dissipation, Γ_m , and the radiation-pressure-induced damping resulting from scattering photons to the upper/lower sideband, $\Gamma = \Gamma_+ - \Gamma_-$, where $\Gamma_{\pm} = 4g^2\kappa/[\kappa^2 + 4(\Delta \pm \Omega_m)^2]$. Here g is the coupling rate between the cavity and the mechanical mode, which depends on the amplitude of the drive: $g = Gx_{\text{zp}}\sqrt{n_d}$. Figure 2c shows the measured values of κ , g and Γ'_m as the drive increases. The radiation-pressure damping of the mechanical oscillator becomes pronounced above a cavity drive amplitude of approximately 75 photons, at which point $\Gamma'_m = 2\Gamma_m$ and the mechanical linewidth has doubled. Note that the increased damping rate can be switched off at any time by removing the cooling drive, returning the mechanical oscillator to its intrinsic quality factor, Q_m .

Whereas the absolute value of the displacement imprecision decreases with increasing power, the visibility of the thermal mechanical peak no longer improves once the radiation-pressure force becomes the dominant dissipation mechanism for the membrane. By expressing the imprecision as equivalent thermal quanta of the oscillator, $n_{\text{imp}} = \Gamma'_m S_x^{\text{imp}} / 8x_{\text{zp}}^2$, we see that the visibility of the thermal noise above the imprecision no longer improves once the drive is much greater than $n_d \approx 100$ (Fig. 2d). This is because a linear decrease in S_x^{imp} is balanced by a linear increase in Γ'_m due to radiation-pressure damping. The asymptotic value of n_{imp} is a direct measure of the efficiency of the microwave measurement. Ideally, for a lossless circuit, a quantum-limited microwave measurement would imply $n_{\text{imp}} = 1/4$. The incorporation of the low-noise Josephson parametric amplifier improves n_{imp} close to this ideal limit, reducing the asymptotic value of n_{imp} from 70 to 1.9 quanta. This level of sensitivity is crucial for resolving any residual thermal motion when cooling into the quantum regime.

Beginning from a cryostat temperature of 20 mK and a thermal occupation of $n_m^T = 40$ quanta, the fundamental mechanical mode of the membrane is cooled by radiation-pressure forces. Figure 3a shows the displacement spectral density of the motional sideband as n_d is increased from 18 to 4,500 photons, along with fits to a Lorentzian lineshape (shaded areas). As described above, this increased drive results in three effects on the spectra: lower noise floor, wider resonances and smaller shaded area. Because the shaded area corresponds to the mean-square membrane displacement, it directly measures the effective temperature of the mode. At a drive intensity with 4,000

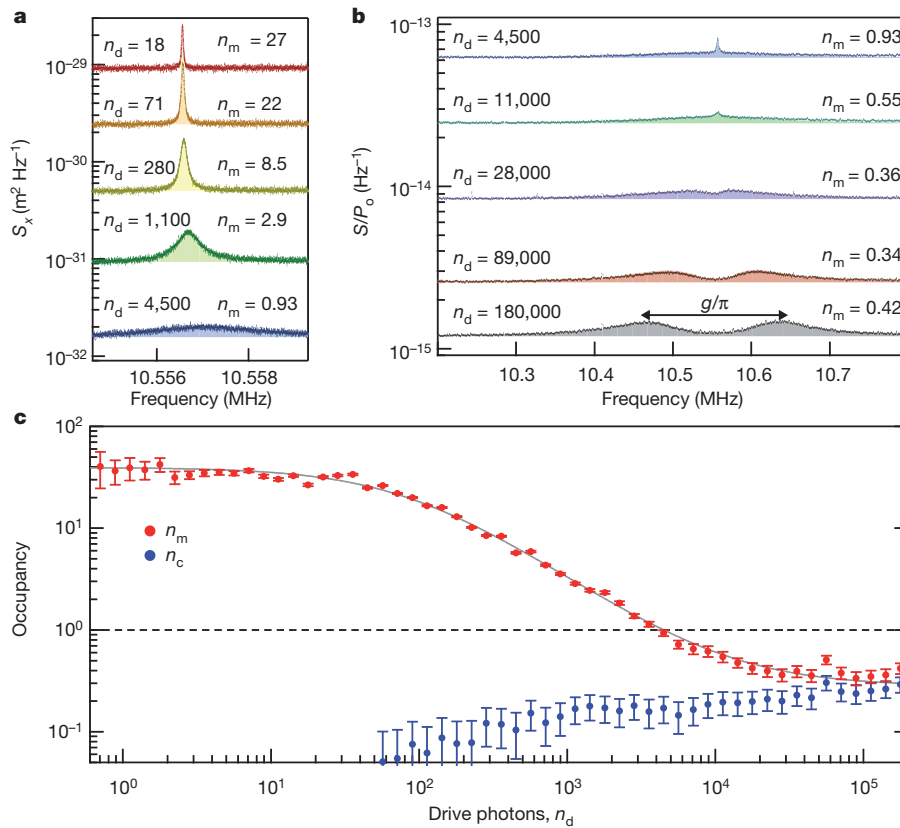


Figure 3 | Sideband cooling the mechanical mode to the ground state. **a**, The displacement noise spectra and Lorentzian fits (shaded regions) for five different drive powers. With higher power, the mechanical mode is both damped (larger linewidth) and cooled (smaller area) by the radiation pressure forces. **b**, Over a broader frequency span, the normalized sideband noise spectra clearly show both the narrow mechanical peak and a broader cavity peak due to finite occupancy of the mechanical and electrical modes, respectively. A small, but resolvable, thermal population of the cavity appears as the drive power increases, setting the limit for the final occupancy of the coupled

optomechanical system. At the highest drive power, the coupling rate between the mechanical oscillator and the microwave cavity exceeds the intrinsic dissipation of either mode, and the system hybridizes into optomechanical normal modes. **c**, Starting in thermal equilibrium with the cryostat at $T = 20$ mK, sideband cooling reduces the thermal occupancy of the mechanical mode from $n_m = 40$ into the quantum regime, reaching a minimum of $n_m = 0.34 \pm 0.05$. Error bars, s.d. These data demonstrate that the parametric interaction between photons and phonons can initialize the strongly coupled, electromechanical system in its quantum ground state.

photons in the cavity, the thermal occupation is reduced below one quantum of mechanical motion, entering the quantum regime.

Observing the noise spectrum over a broader frequency range reveals that in addition to the mechanical Lorentzian peak with linewidth Γ'_m , there is also a Lorentzian peak with linewidth κ whose area corresponds to the finite thermal occupation n_c of the cavity. Over this range, it is no longer valid to evaluate the cavity parameters at a single frequency to infer the spectrum in units of S_x . Instead, Fig. 3b shows the noise spectrum in units of sideband power normalized by the power at the drive frequency, S/P_o . These two sources of noise, originating from either the mechanical or the electrical mode, interfere with each other and result in noise squashing¹⁴ and eventually normal-mode splitting²⁸ once $2g > \kappa/\sqrt{2}$. Using a quantum-mechanical description applied to our circuit^{14,24}, the expected noise spectrum as a function of frequency ω is:

$$S = \hbar\omega \left(\frac{1}{2} + n_{\text{add}} + 2\kappa_{\text{ex}} \frac{\kappa n_c (\Gamma_m^2 + 4\delta^2) + 4\Gamma_m n_m^T g^2}{|4g^2 + (\kappa + 2j(\delta + \tilde{\Delta}))(\Gamma_m + 2j\delta)|^2} \right) \quad (1)$$

where $j = \sqrt{-1}$, $\delta = \omega - \Omega_m$, $\tilde{\Delta} = \omega_d + \Omega_m - \omega_c$, and n_{add} is added noise of the microwave measurement expressed as an equivalent number of microwave photons. Figure 3b shows the measured spectra and corresponding fits (shaded regions) to equation (1) as the electromechanical system evolves first into the quantum regime ($n_m, n_c < 1$) and then into the strong-coupling regime ($2g > \kappa/2$). The results are summarized in Fig. 3c, where the thermal occupancy of both the mechanical and electrical modes is shown as a function of n_d . For low drive power, the cavity shows no resolvable thermal population (to within our measurement uncertainty of 0.05 quanta), as expected for a 7.5-GHz mode at 20 mK. Although it is unclear whether the observed population at higher drive power is a consequence of direct heating of the substrate, heating of the microwave attenuators preceding the circuit, or intrinsic cavity frequency noise, we have determined that it is not the result of frequency or amplitude noise of our microwave generator, as this noise is reduced far below the microwave shot-noise level with narrow-band filtering and cryogenic attenuation (see Supplementary Information). Sideband cooling can never reduce the occupancy of the mechanical mode below that of the cavity. Therefore, in order for the system to access the quantum regime, the thermal population of the cavity must remain less than one quantum. Assuming $\Omega_m \gg \kappa$, the final occupancy of a mechanical mode is²⁸:

$$n_m = n_m^T \left(\frac{\Gamma_m}{\kappa} \frac{4g^2 + \kappa^2}{4g^2 + \kappa\Gamma_m} \right) + n_c \left(\frac{4g^2}{4g^2 + \kappa\Gamma_m} \right) \quad (2)$$

This equation shows that for moderate coupling ($\sqrt{\kappa\Gamma_m} \ll g \ll \kappa$) the cooling of the mechanical mode is linear in the number of drive photons. Beyond this regime, the onset of normal-mode splitting abates further cooling. Here the mechanical cooling rate becomes limited not by the coupling between the mechanical mode and the cavity, but instead by the coupling rate κ between the cavity and its environment²⁸. Thus, the final occupancy of the mechanical mode can never be reduced to lower than $n_m^T \Gamma_m / \kappa$, and a stronger parametric drive will only increase the Rabi frequency at which the thermal excitations oscillate between the cavity and mechanical modes. For our device, as the coupling is increased, we first cool to the ground state and then enter the strong-coupling regime ($n_m^T \Gamma_m < \kappa < g$). Once n_d exceeds 2×10^4 , the mechanical occupancy converges towards the cavity population, reaching a minimum of 0.34 ± 0.05 quanta. At the highest power drive ($n_d = 2 \times 10^5$), the mechanical mode has hybridized with the cavity, resulting in the normal-mode splitting characteristic of the strong-coupling regime^{16,29}. This level of coupling is required to use the hybrid system for quantum information processing. The strong-coupling regime ensures that quantum states of this combined system may be manipulated faster than they decohere from spurious interaction with either electromagnetic or mechanical environments.

Taken together, the measurements shown in Figs 2 and 3 quantify the overall measurement efficiency of the system. The Heisenberg limit

requires that a continuous displacement measurement is necessarily accompanied by a back-action force^{3,13,24}, such that $\sqrt{S_x^{\text{imp}} S_F^{\text{ba}}} \geq \hbar$, where S_F^{ba} is the force noise spectral density from back-action alone. From the thermal occupancy and damping rate of the mechanical mode, we extract a total force spectral density $S_F^{\text{tot}} = 4\hbar\Omega_m m \Gamma'_m (n_m + 1/2)$. By attributing all of S_F^{tot} to back-action, we can place a conservative upper bound on the imprecision-back-action product with $\sqrt{S_x^{\text{imp}} S_F^{\text{tot}}}/\hbar = \sqrt{n_{\text{imp}}(n_m + 1/2)} \leq 5.1 \pm 0.4$. The minimum achievable value with our detuned probe tone (see Supplementary Information) is $\sqrt{S_x^{\text{imp}} S_F^{\text{ba}}} = \hbar\sqrt{2}$, making this experiment a factor of 3.6 away from the Heisenberg-limit for displacement detection, the narrowest gap achieved to date^{15,24}.

Looking forward, this technology offers a feasible route to achieve many of the long-standing goals for mechanical quantum systems. Whereas the resolved-sideband regime is well-suited for efficient sideband cooling, it makes a simultaneous measurement of both the upper and lower sideband difficult. By simply increasing the bandwidth of the cavity, future experiments could feasibly measure the zero-point motion of the mechanical mode, as well as observe the fundamental asymmetry between the rates of emission and absorption of phonons¹. Other prospects include quantum non-demolition measurements³ and the generation of entangled states of mechanical motion^{17,18}. Furthermore, combining this device with a superconducting qubit²¹ would allow for the direct measurement of the mechanical oscillator's energy states and the preparation of arbitrary quantum states of mechanical motion³⁰. Because the interaction between the 10.6-MHz mechanical mode and the 7.5-GHz microwave cavity is parametric, the coupling strength is inherently tunable, and can be turned on and off quickly. Thus, once a quantum state is transferred into the mechanical mode, it can be stored there for a time $\tau_{\text{th}} = 1/(n_m^T \Gamma_m) > 100 \mu\text{s}$ before absorbing one thermal phonon from its environment. As this timescale is much longer than typical coherence times of superconducting qubits, mechanical modes offer the potential for delay and storage of quantum information. Lastly, mechanical objects provide a generic system for interacting with a wide range of different physical systems—ranging from magnetic spins to optical photons—leading towards future methods for engineering the coherent transfer of quantum information between vastly different forms of quanta²⁰.

The power and versatility of sideband cooling techniques have now been used to bring a high quality, macroscopic ($\sim 10^{12}$ atoms) low-frequency mechanical oscillator into the quantum regime. This electromechanical system simultaneously demonstrates ground-state preparation, strong-coupling and near quantum-limited position detection, paving the way to accessing the quantum nature of long-lived motional states.

Received 22 February; accepted 2 June 2011.

Published online 6 July 2011.

1. Diedrich, F., Bergquist, J. C., Itano, W. M. & Wineland, D. J. Laser cooling to the zero-point energy of motion. *Phys. Rev. Lett.* **62**, 403–406 (1989).
2. Anderson, M. H., Ensher, J. R., Matthews, M. R., Wieman, C. E. & Cornell, E. A. Observation of Bose-Einstein condensation in a dilute atomic vapor. *Science* **269**, 198–201 (1995).
3. Braginsky, V. B. & Khalili, F. Y. *Quantum Measurement* (Cambridge Univ. Press, 1992).
4. Kippenberg, T. J. & Vahala, K. J. Cavity optomechanics: back-action at the mesoscale. *Science* **321**, 1172–1176 (2008).
5. Braginsky, V. B., Manukin, A. B. & Tikhonov, M. Y. Investigation of dissipative ponderomotive effects of electromagnetic radiation. *Sov. Phys. JETP* **31**, 829–830 (1970).
6. Blair, D. G. *et al.* High sensitivity gravitational wave antenna with parametric transducer readout. *Phys. Rev. Lett.* **74**, 1908–1911 (1995).
7. Cuthbertson, B. D., Tobar, M. E., Ivanov, E. N. & Blair, D. G. Parametric back-action effects in a high-Q cryogenic sapphire transducer. *Rev. Sci. Instrum.* **67**, 2435–2442 (1996).
8. Teufel, J. D., Harlow, J. W., Regal, C. A. & Lehnert, K. W. Dynamical backaction of microwave fields on a nanomechanical oscillator. *Phys. Rev. Lett.* **101**, 197203 (2008).

9. Thompson, J. D. *et al.* Strong dispersive coupling of a high-finesse cavity to a micromechanical membrane. *Nature* **452**, 72–75 (2008).
10. Gröblacher, S. *et al.* Demonstration of an ultracold micro-optomechanical oscillator in a cryogenic cavity. *Nature Phys.* **5**, 485–488 (2009).
11. Park, Y.-S. & Wang, H. Resolved-sideband and cryogenic cooling of an optomechanical resonator. *Nature Phys.* **5**, 489–493 (2009).
12. Lin, Q., Rosenberg, J., Jiang, X., Vahala, K. J. & Painter, O. Mechanical oscillation and cooling actuated by the optical gradient force. *Phys. Rev. Lett.* **103**, 103601 (2009).
13. Schliesser, A., Arcizet, O., Rivière, R., Anetsberger, G. & Kippenberg, T. J. Resolved-sideband cooling and position measurement of a micromechanical oscillator close to the Heisenberg uncertainty limit. *Nature Phys.* **5**, 509–514 (2009).
14. Rocheleau, T. *et al.* Preparation and detection of a mechanical resonator near the ground state of motion. *Nature* **463**, 72–75 (2010).
15. Rivière, R. *et al.* Optomechanical sideband cooling of a micromechanical oscillator close to the quantum ground state. Preprint at (<http://arXiv.org/abs/1011.0290>) (2010).
16. Teufel, J. D. *et al.* Circuit cavity electromechanics in the strong coupling regime. *Nature* **471**, 204–208 (2011).
17. Bose, S., Jacobs, K. & Knight, P. L. Preparation of nonclassical states in cavities with a moving mirror. *Phys. Rev. A* **56**, 4175–4186 (1997).
18. Mancini, S., Man'ko, V. I. & Tombesi, P. Ponderomotive control of quantum macroscopic coherence. *Phys. Rev. A* **55**, 3042–3050 (1997).
19. Marshall, W., Simon, C., Penrose, R. & Bouwmeester, D. Towards quantum superpositions of a mirror. *Phys. Rev. Lett.* **91**, 130401 (2003).
20. Regal, C. A. & Lehnert, K. W. From cavity electromechanics to cavity optomechanics. *J. Phys. Conf. Ser.* **264**, 012025 (2011).
21. O'Connell, A. D. *et al.* Quantum ground state and single-phonon control of a mechanical resonator. *Nature* **464**, 697–703 (2010).
22. Marquardt, F., Chen, J. P., Clerk, A. A. & Girvin, S. M. Quantum theory of cavity-assisted sideband cooling of mechanical motion. *Phys. Rev. Lett.* **99**, 093902 (2007).
23. Wilson-Rae, I., Nooshi, N., Zwerger, W. & Kippenberg, T. J. Theory of ground state cooling of a mechanical oscillator using dynamical backaction. *Phys. Rev. Lett.* **99**, 093901 (2007).
24. Clerk, A. A., Devoret, M. H., Girvin, S. M., Marquardt, F. & Schoelkopf, R. J. Introduction to quantum noise, measurement, and amplification. *Rev. Mod. Phys.* **82**, 1155–1208 (2010).
25. Cicak, K. *et al.* Low-loss superconducting resonant circuits using vacuum-gap-based microwave components. *Appl. Phys. Lett.* **96**, 093502 (2010).
26. Castellanos-Beltran, M. A., Irwin, K. D., Hilton, G. C., Vale, L. R. & Lehnert, K. W. Amplification and squeezing of quantum noise with a tunable Josephson metamaterial. *Nature Phys.* **4**, 929–931 (2008).
27. Teufel, J. D., Donner, T., Castellanos-Beltran, M. A., Harlow, J. W. & Lehnert, K. W. Nanomechanical motion measured with an imprecision below that at the standard quantum limit. *Nature Nanotechnol.* **4**, 820–823 (2009).
28. Dobrindt, J. M., Wilson-Rae, I. & Kippenberg, T. J. Parametric normal-mode splitting in cavity optomechanics. *Phys. Rev. Lett.* **101**, 263602 (2008).
29. Gröblacher, S., Hammerer, K., Vanner, M. R. & Aspelmeyer, M. Observation of strong coupling between a micromechanical resonator and an optical cavity field. *Nature* **460**, 724–727 (2009).
30. Hofheinz, M. *et al.* Synthesizing arbitrary quantum states in a superconducting resonator. *Nature* **459**, 546–549 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. W. Sanders for taking the micrograph in Fig. 1 and the JILA instrument shop for fabrication and design of the cavity filter. This work was supported by NIST and the DARPA QuASAR programme. T.D. acknowledges support from the Deutsche Forschungsgemeinschaft (DFG). This Letter is a contribution of the US government and not subject to copyright.

Author Contributions J.D.T. and R.W.S. conceived the device. J.D.T. designed the circuit. J.D.T. and D.L. fabricated the devices. J.D.T. and T.D. set up the experiment, performed the measurements and analysed the data. J.D.T., T.D., R.W.S. and K.W.L. discussed the results and wrote the manuscript. All authors provided experimental support and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.D.T. (john.teufel@nist.gov).

ERRATUM

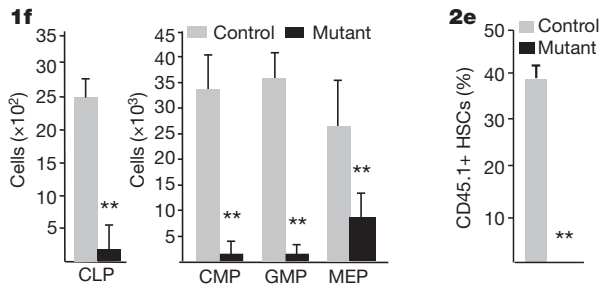
doi:10.1038/nature10270

The Lkb1 metabolic sensor maintains haematopoietic stem cell survival

Sushma Gurumurthy, Stephanie Z. Xie, Brinda Alagesan, Judith Kim, Rushdia Z. Yusuf, Borja Saez, Alexandros Tzatsos, Fatih Ozsolak, Patrice Milos, Francesco Ferrari, Peter J. Park, Orian S. Shirihai, David T. Scadden & Nabeel Bardeesy

Nature **468**, 659–663 (2010)

In this Article, Figs 1f and 2e appeared incorrectly (the corrected figure parts are shown below). In Fig. 1f, the y-axis of the CLP graph should read “Cells ($\times 10^2$)” and the y-axis of the CMP graph should read “Cells ($\times 10^3$)”. In Fig. 2e, the control bar should be grey, not black.



Direct generation of functional dopaminergic neurons from mouse and human fibroblasts

Massimiliano Caiazzo¹, Maria Teresa Dell'Anno^{1*}, Elena Dvoretzkova^{2*}, Dejan Lazarevic^{3,4}, Stefano Taverna², Damiana Leo², Tatyana D. Sotnikova², Andrea Menegon⁵, Paola Roncaglia⁴, Giorgia Colciago¹, Giovanni Russo², Piero Carninci⁶, Gianni Pezzoli⁷, Raul R. Gainetdinov², Stefano Gustincich^{4,8}, Alexander Dityatev² & Vania Broccoli¹

Transplantation of dopaminergic neurons can potentially improve the clinical outcome of Parkinson's disease, a neurological disorder resulting from degeneration of mesencephalic dopaminergic neurons^{1,2}. In particular, transplantation of embryonic-stem-cell-derived dopaminergic neurons has been shown to be efficient in restoring motor symptoms in conditions of dopamine deficiency^{3,4}. However, the use of pluripotent-derived cells might lead to the development of tumours if not properly controlled⁵. Here we identified a minimal set of three transcription factors—*Mash1* (also known as *Ascl1*), *Nurr1* (also known as *Nr4a2*) and *Lmx1a*—that are able to generate directly functional dopaminergic neurons from mouse and human fibroblasts without reverting to a progenitor cell stage. Induced dopaminergic (iDA) cells release dopamine and show spontaneous electrical activity organized in regular spikes consistent with the pacemaker activity featured by brain dopaminergic neurons. The three factors were able to elicit dopaminergic neuronal conversion in prenatal and adult fibroblasts from healthy donors and Parkinson's disease patients. Direct generation of iDA cells from somatic cells might have significant implications for understanding critical processes for neuronal development, *in vitro* disease modelling and cell replacement therapies.

Seminal studies have demonstrated that functional neurons can be generated independently of stem cells by direct cell conversion through genetics-based approaches⁶. More recently, in a set of elegant experiments, fibroblasts have been directly converted into neuronal cells (iNs) by the forced expression of the three neurodevelopmental factors *Mash1*, *Brn2* (also known as *Pou3f2*) and *Myt1l*⁷. However, iNs represent a heterogeneous population of glutamatergic and GABAergic neurons and their degree of global reprogramming remains to be properly characterized. It is thus unclear whether a specific neuronal subtype can be preferentially induced from direct reprogramming of heterologous cells. Therefore, we aimed to generate dopaminergic neurons through the direct conversion of somatic cells by forced expression of lineage-specific factors that act during brain development^{8,9}. Initially, we transduced mouse embryonic fibroblasts (MEFs) from TH-GFP transgenic mice¹⁰ with a mixture of doxycycline (dox)-inducible lentiviruses expressing all selected factors (11 dopaminergic and 3iN; Supplementary Table 1) or with DsRed retrovirus (negative control) (Fig. 1a–d). We did not observe any GFP⁺ cells in MEFs 10 days after Ds-Red retrovirus infection or in culture without any viral infection (Fig. 1d, inset). In contrast, transduction of all factors resulted in the generation of a small number of bright GFP⁺ cells ($1.8 \pm 0.8\%$) (Supplementary Fig. 1d–f). We next sought to determine the minimal set of genes required for dopaminergic neuronal induction. Given its essential role as a proneural gene during neurogenesis, *Mash1* was

introduced into MEFs together with each other single dopaminergic factor. Reporter gene expression was elicited only when *Mash1* was combined with *Nurr1*, a critical determinant of dopaminergic neuronal specification and survival during development and in adulthood¹¹. However, *Mash1/Nurr1* combined activation elicited a modest increase of GFP⁺ cells ($8 \pm 2\%$) (Supplementary Fig. 1g–i). Therefore, we added a third molecule of the 12 remaining and scored for the rate and morphology of GFP⁺ cells in each combination. Surprisingly, only *Lmx1a* and in part *Lmx1b* ($18 \pm 3\%$ versus $13 \pm 3\%$ of GFP⁺ cells, respectively) were able to synergize with *Mash1/Nurr1*, robustly increasing the generation of GFP⁺ cells with an evidently complex neuronal morphology (Fig. 1h and Supplementary Fig. 1s–y). Using the *Mash1/*

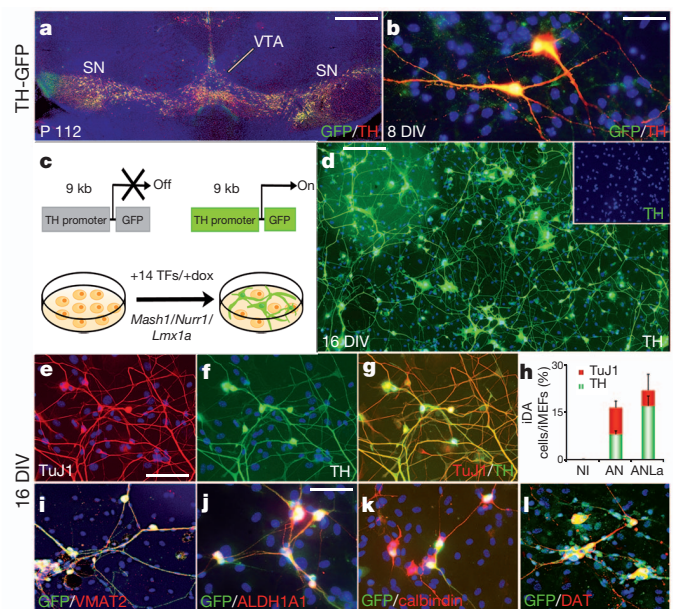


Figure 1 | *Mash1*, *Nurr1* and *Lmx1a* reprogram mouse fibroblasts into iDA cells. **a**, **b**, TH and GFP detection in TH-GFP adult brain (**a**) and ventral midbrain primary cell culture (**b**). DIV, days *in vitro*. SN, substantia nigra; VTA, ventral tegmental area. P112, postnatal day 112. **c**, Scheme of dopaminergic transcription factors (TFs) screening. **d**, TH staining in iDA cells and in uninfected MEFs (inset) after 16 days *in vitro*. **e–g**, iDA cells are positive for the dopaminergic markers TH (**e–g**), VMAT2, ALDH1A1, calbindin and DAT (**i–l**). **h**, Quantification of TuJ1⁺ and TH⁺ cells. AN, *Mash1*, *Nurr1*; ANLa, *Mash1*, *Nurr1*, *Lmx1a*; NI, non-induced. Data are presented as mean \pm s.e.m. Scale bars: 500 μ m (**a**), 20 μ m (**b**, **j**), 50 μ m (**e–g**, **k**, **l**) and 100 μ m (**d**, **i**).

¹Stem Cells and Neurogenesis Unit, Division of Neuroscience, San Raffaele Scientific Institute, 20132 Milan, Italy. ²Department of Neuroscience and Brain Technologies, Istituto Italiano di Tecnologia, via Morego, 30, 16163 Genoa, Italy. ³CBM Srl. Area Science park, Basovizza, SS14, km165, 34149 Trieste, Italy. ⁴Sector of Neurobiology, International School for Advanced Studies (SISSA), via Bonomea, 265, 34136 Trieste, Italy. ⁵Advanced Light and Electron Microscopy Bio-Imaging Centre, Experimental Imaging Centre, San Raffaele Scientific Institute, 20132 Milan, Italy. ⁶Omics Science Center, RIKEN Yokohama Institute, 1-7-22 Suehiro-chō, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan. ⁷Parkinson Institute, Istituti Clinici di Perfezionamento, 20126 Milan, Italy. ⁸The Giovanni Armenise-Harvard Foundation Laboratory, 34136 Trieste, Italy.

*These authors contributed equally to this work.

Nurr1/Lmx1a factor combination, double GFP⁺/TH⁺ cells represented the majority of the induced TuJ1 neuronal cells (85 ± 4%). Supplementation of a fourth factor among the remaining ones failed to produce any further increase in GFP⁺ cells, with *Brn2* and *Myt1l*, the other two iN factors, even reducing the overall reprogramming efficiency (data not shown). For these reasons we focused on cells reprogrammed exclusively with the *Mash1/Nurr1/Lmx1a* factor combination. The same gene cocktail was also proficient in reprogramming adult mouse fibroblasts with high efficiency (Supplementary Fig. 2).

Sixteen days after reprogramming, a large number of GFP⁺ cells expressed many of the distinctive components of the dopaminergic machinery, such as TH, vesicular monoamine transporter 2 (VMAT2; also known as SLC18A2), dopamine transporter (DAT; also known as SLC6A3), as well as aldehyde dehydrogenase 1a1 (ALDH1A1) and calbindin (Fig. 1e–l). Conversely, markers associated with adrenergic (dopamine-β-hydroxylase (DBH)) or serotonergic (tryptophan hydroxylase 1 or 2 (TPH1/2)) serotonin transporter (SERT; also known as SLC6A4) neurons were not induced (data not shown). Transcriptional analysis by reverse transcription–polymerase chain reaction (RT–PCR) confirmed the activation of the dopamine-specific gene network including the endogenous expression of *Nurr1* and *Lmx1a* (Supplementary Fig. 3). Global expression analysis showed that iDA cells clustered with A9 and A10 adult mesencephalic dopaminergic (mDA) neurons rather than with their fibroblasts of origin as illustrated by hierarchical clustering (Fig. 2a, b) and the general degree of gene expression overlap (Fig. 2c). Of note, many representative genes of the dopaminergic phenotype, such as *Th*, *Vmat2*, *Aadc* (also known as *Ddc*), *Ret*, *Gfra1*, *Foxa1*, *Gdnf* and *Drd2* were highly enriched (Fig. 2d). Conversely, genes coding for adrenergic and serotonergic biosynthetic enzymes were found to be not upregulated in the reprogrammed cells (Fig. 2e). Moreover, the fibroblast markers *Twist2*, *Zeb2*, *Tgfb1l1* and *Chd2* (ref. 12) were downregulated in iDA cells (Fig. 2f). These findings indicate that the genetic reprogramming erased the majority of the evident expression hallmarks of the cell of origin, while specifically inducing the dopaminergic neuronal phenotype and not those of other

closely related neuronal subtypes. It should be noted that iDA expression profiling was close but distinguishable from that of mDA neurons, with 160 genes differently expressed with a ≥5-fold change (Supplementary Table 3). Whether this might indicate the presence of residual fibroblast gene expression in iDA cells remains to be addressed.

Interestingly, *Th* and *Vmat2* promoter regions were highly demethylated in dopaminergic neuronal cells whereas they were fully methylated in parental fibroblasts, indicating their epigenetic reactivation during dopaminergic neuronal conversion (Supplementary Fig. 4).

GFP⁺ cells induced by the three factors showed an elaborate neuronal morphology with multiple and long processes (Fig. 1d–l). Hence, we asked whether induced neuronal cells establish synaptic contacts in culture. Notably, synaptic resident proteins such as synaptotagmin I (SYT1) and synapsin (SYN) were localized in discrete puncta and colocalized with TH immunolabelling, suggesting the establishment of dopaminergic synaptic terminals (Supplementary Fig. 5). Moreover, successful FM4-64 dye uptake at TH⁺ synaptic boutons indicated active synaptic processes (Supplementary Fig. 5).

Next, we performed patch-clamp recordings of GFP⁺ iDA cells (*n* = 16) as well as primary mDA neurons (*n* = 12) to compare their respective physiological properties^{13,14}. iDA cells had higher cell resistance and lower capacitance than primary dopaminergic neurons, but showed normal resting membrane potential, normal Na⁺ currents (Fig. 3a), overshooting action potentials (Fig. 3c), and even more prominent K⁺ currents (Fig. 3b) and afterspike hyperpolarizations (Supplementary Table 4). More than 80% of iDA cells showed rhythmic discharges (Fig. 3d, e) at an average frequency of 2.6 Hz. The identity of voltage-gated inward Na⁺ and outward K⁺ currents in iDA cells has been verified pharmacologically (Supplementary Fig. 6). Next, we noted that iDA cells, like mDA neurons, express high levels of the D2 receptor (Fig. 3f). To verify whether the dopamine receptors were functional, we applied the specific D2/D3 receptor agonist quinpirole (1 μM), which markedly suppressed neuronal firing in 6 of 10 recorded cells in a reversible manner (Fig. 3g, h). Next, we used amperometry for real-time electrochemical detection of monoamine secretion from iDA cells^{15,16}. When carbon fibre electrodes were placed adjacent to GFP⁺ cells (Fig. 3i), depolarization of 4 cells with 25 mM K⁺ resulted in numerous amperometric events (Fig. 3i), reflecting quantal secretion of monoamines. Furthermore, high-performance liquid chromatography (HPLC) measurements revealed that iDA cells contain high levels of intracellular dopamine, detectable in pellet preparations, which can be released in the extracellular medium on stimulation with 50 mM KCl (Fig. 3j). Thus, reprogrammed cells show several major properties of dopaminergic neurons in terms of spontaneous spiking activity, temporal parameters of action potentials, inhibition of cell firing through D2 autoreceptors and controlled dopamine release.

Next, to determine the temporal requirements for the three exogenous factors to induce a stable reprogrammed cell state, infected MEFs were treated with dox for different time windows, after which dox was withdrawn. Only when fibroblasts were treated with dox for 6 or more days were numerous neuronal cells, mostly TH⁺, observed (Supplementary Fig. 7). Thus, reprogramming is a relatively rapid process that requires the expression of the three factors for only 6 days. At the same time, iDA cells achieved a stable neuronal state over time independently from viral transgene expression and even at 18 days of dox withdrawal, iDA cells were found at the same number, and exhibited similar spontaneous firing as iDA cells constantly cultivated in the presence of dox (Supplementary Fig. 8 and Supplementary Table 4).

Reprogramming of fibroblasts into differentiated neuronal cells might occur directly or by passing first through neural progenitors. When the DNA-base analogue BrdU was added from day 2 onwards to label the proliferating cells, virtually all neuronal cells were already post-mitotic after this time (Supplementary Fig. 9b, c, h). Despite the fact that during the first 2 days infected cells were actively proliferating in serum-containing medium, none showed expression of the neural progenitor molecular markers *Sox2*, *Ngn2*, *Otx2*, *Lmx1b* and

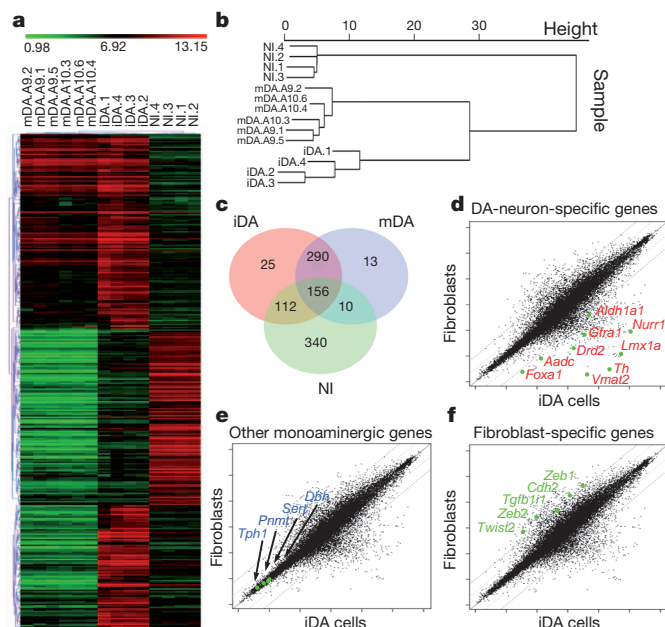


Figure 2 | Mouse iDA cells expression profiling. **a**, Heat-map of genes differentially expressed in RNA-microarray analysis performed on MEFs (NI), iDA cells and brain mDAs (A9–A10). **b**, **c**, Hierarchical clustering (**b**) and general degree of overlapping expression (**c**) among the three cell populations analysed. **d–f**, Scatter plots show that in iDA cells the majority of dopaminergic (DA) markers are increased, whereas other monoaminergic neuronal markers are not activated and fibroblasts markers are silenced.

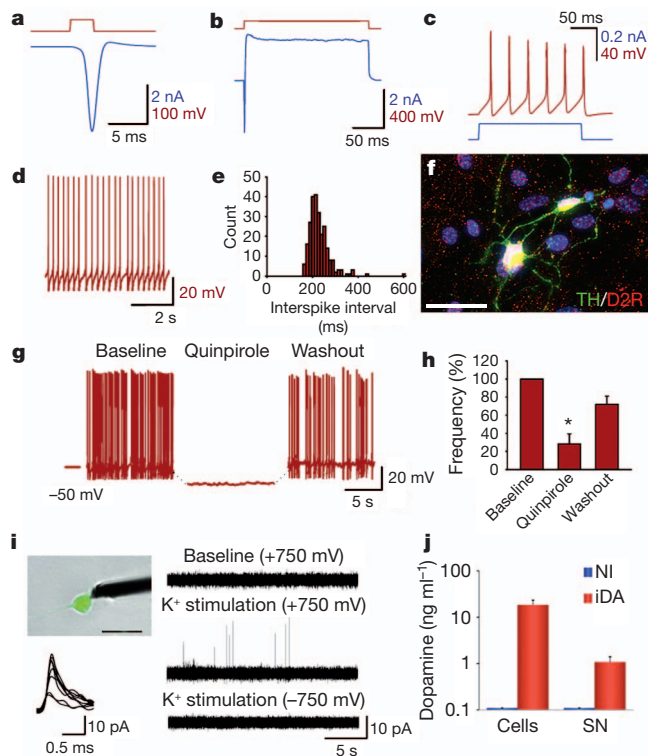


Figure 3 | Functional characterization of mouse iDA cells. **a, b,** Whole-cell voltage-clamp recording of Na^+ and K^+ currents. **c,** Current-clamp recording of multiple action potentials evoked by current injection. **d, e,** Current-clamp recording and interspike interval frequency of spontaneous action potentials. **f,** D2 receptor (D2R) staining. **g, h,** Effect of the D2/D3 agonist quinpirole on spiking frequency (**g**) and its statistical analysis (**h**) (* $P = 0.005$, paired t -test, $n = 6$). **i,** Amperometric recordings of monoamine release after K^+ stimulation; release events are shown with high-resolution below the image of the recorded cell. **j,** Dopamine content measured by HPLC in uninfected (NI) and iDA cells, both in cell pellets and in the supernatant (SN) after K^+ stimulation. All cells are 16 days *in vitro* unless otherwise stated. Scale bars: 50 μm (**b**) and 20 μm (**i**). Data are presented as mean \pm s.e.m.

En1 (ref. 17) (Supplementary Fig. 9i). Furthermore, we used a genetic tracing system based on activation of the *Sox2* ^{β -geo} LacZ reporter¹⁸ showing that, as a proof-of-principle, LacZ activity was easily visualized upon reprogramming of *Sox2* ^{β -geo} fibroblasts into induced pluripotent stem cells. By contrast, the reporter was never activated from the same cells when engaged into direct iDA reprogramming (Supplementary Fig. 9j). Altogether, these findings are inconsistent with the occurrence of detectable cell intermediates during the reprogramming of fibroblasts into iDA cells.

Next, the *in vivo* differentiation potential of iDA cells was assessed by orthotopic transplantations into neonatal mouse brains. Four days after viral transgene induction, infected cells were grafted into the ventricle of mouse newborn brains. Two and six weeks after transplantations, GFP⁺ cells were found integrated in the host tissue showing an extremely elaborated morphology (Supplementary Figs 10 and 11). Most of the GFP⁺ grafted neuronal cells were positive for TH, AADC, VMAT2 and DAT, indicating the achievement of a full neuronal dopaminergic cell fate (Supplementary Fig. 11b–g, i–l). Injection of brief supra-threshold current pulses evoked overshooting action potentials, and large Na^+ and K^+ currents were activated by depolarizing voltage steps (Supplementary Fig. 11o, p). Therefore, iDA cells maintain excitability and major currents *in vivo* even after an extensive period of time from grafting.

We then translated the same procedure to the human system by initially infecting IMR90 fetal fibroblasts. After 18 days from infection, we scored numerous TuJ1⁺ and TH⁺ neuronal cells accounting,

respectively, for $10 \pm 4\%$ and $6 \pm 2\%$ of the infected cells (Supplementary Fig. 12a–c). We then reprogrammed adult human fibroblasts from two healthy donors (aged 42 and 55) and from two patients with genetic forms of Parkinson's disease (Supplementary Table 5). Both healthy and diseased adult cells showed a comparable propensity to convert into neuronal cells, accounting for an estimated efficiency for TuJ1⁺ and TH⁺ cells of $5 \pm 1\%$ and $3 \pm 1\%$, respectively (Fig. 4a–g). iDA cells were positive for ALDH1A1, TH, AADC, VMAT2 and DAT by immunocytochemistry (Fig. 4a–f) and gene expression analysis (Supplementary Fig. 13). Cell conversion was stable over time as the number and morphology of human iDA cells was not obviously affected up to day 24 after reprogramming even when dox was withdrawn from day 6 onwards (Fig. 4h, i).

Recordings in five infected fetal human iDA cells showed that the electrophysiological properties of these cells resemble mouse iDA cells (Supplementary Fig. 12). Recordings in eight infected adult human iDA cells revealed less mature phenotypes (Fig. 4j–l), with mean amplitudes: 0.7 ± 0.1 nA for Na^+ currents; 1.2 ± 0.1 nA for delayed rectifier K^+ currents; 78 ± 3 mV for action potentials; and 7.5 ± 2 mV for afterdepolarization. The identity of Na^+ and K^+ currents was confirmed pharmacologically (Fig. 4m, n). Most importantly, depolarization of

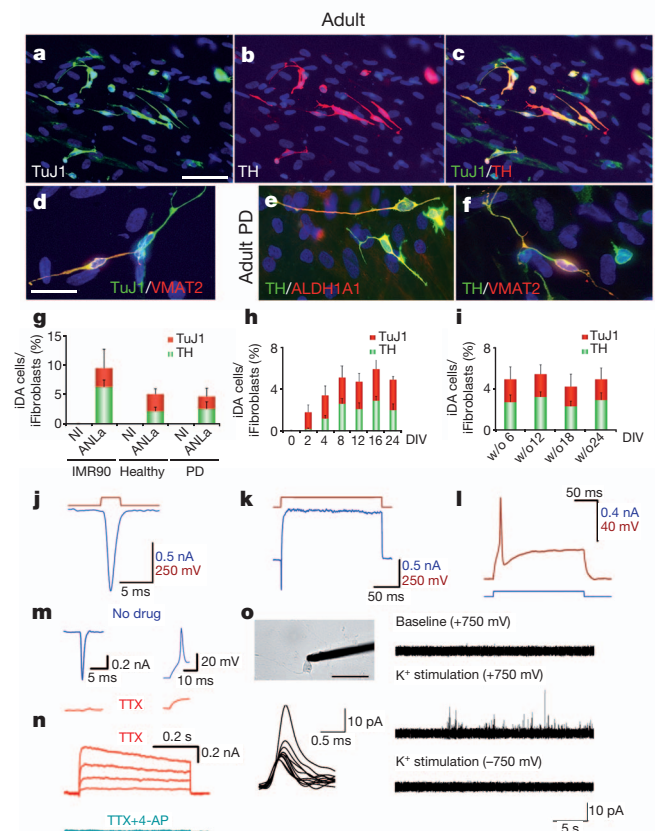


Figure 4 | Characterization of human fibroblasts reprogrammed into iDA cells. **a–f,** Fibroblasts from a healthy donor (**a–d**) and a Parkinson's disease (PD) patient (**e, f**) show a comparable efficiency in dopaminergic neuronal conversion. **g,** Quantification of iDA cells obtained from fetal (IMR90), healthy and Parkinson's disease adult fibroblasts. **h,** Quantification of TuJ1⁺ and TH⁺ cells in a time-course study from 0 to 24 days *in vitro* (DIV). **i,** Quantification of TuJ1⁺ and TH⁺ reprogrammed cells kept with (w) or without (w/o) doxycycline for 6, 12, 18 or 24 DIV. Data are presented as mean \pm s.e.m. **j, k,** Whole-cell voltage-clamp recording of Na^+ and K^+ currents. **l,** Whole-cell current-clamp recording of single action potential elicited by a minimal depolarization. **m, n,** Suppression of Na^+ (**m**) and K^+ (**n**) currents and action potentials by tetrodotoxin (TTX) and 4-AP. **o,** Amperometric recordings of release events after K^+ stimulation; high-resolution pattern is shown below the image of the recorded cell. All cells are 18 days *in vitro* unless otherwise stated. Scale bars: 50 μm (**a–c**) and 20 μm (**d–f, o**).

three cells with 25 mM K⁺ elicited numerous release events detected by amperometric measurements, as described earlier for mouse iDA cells (Fig. 4o). In summary, these experiments indicate that actively spiking dopamine-secreting cells can be induced by forced expression of the three factors in adult human cells from both healthy donors and Parkinson's disease patients.

Here, we demonstrated that the combination of three transcription factors—*Mash1*, *Nurr1* and *Lmx1a*—can rapidly and efficiently induce dopaminergic neuronal cells from mouse and human fibroblasts. Reprogrammed cells are similar to brain dopaminergic neurons in gene expression, and show dopamine release and pacemaker activity that can be modulated via D2 receptors. Importantly, this cell conversion diverges with respect to developmental neuronal lineage commitment as it does not progress through detectable intermediate neuronal stages. How this might impact on the overall functional aspects of the cells will need to be tested further in long-term *in vivo* transplantation studies.

Generation of functional dopaminergic neuronal cells by direct reprogramming opens new possibilities for regenerative therapies for Parkinson's disease and related disorders. However, to be clinically relevant, the overall cell conversion process needs to be highly efficient and the addition of supplementary factors might be helpful¹⁹. Interestingly, a different gene cocktail capable of inducing dopaminergic-like neurons from human fibroblasts has been identified²⁰. This opens the intriguing possibility that different molecular fate determinants reach a similar endpoint even though acting on different transcriptional cascades.

These procedures do not rely on pluripotent stem cells that are prone to tumours in their undifferentiated state. Moreover, the process described here does not pass through proliferative progenitors that also might be tumorigenic²¹. Thus, these procedures might avoid a dangerous drawback of stem cell therapies while providing a sufficient number of functional dopaminergic neurons amenable for autologous cell replacement therapies.

METHODS SUMMARY

Cell culture and viral infection. MEFs were isolated from embryonic day (E)14.5 wild-type or TH-GFP mouse embryos. Adult human fibroblasts isolated from healthy subjects and Parkinson's disease patients as well as from human fetal lung fibroblasts (IMR90) were grown in MEF media. Cells were infected with dox-inducible lentiviruses as previously reported⁷.

Electrophysiology and amperometry. Electrophysiological recordings were performed in on-cell and whole-cell configurations. Carbon-fibre microelectrodes were used for amperometric recordings¹⁵.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 3 March 2011; accepted 15 June 2011.

Published online 3 July 2011.

1. Lindvall, O. & Björklund, A. Cell therapy in Parkinson's disease. *NeuroRx* **1**, 382–393 (2004).
2. Politis, M. *et al.* Serotonergic neurons mediate dyskinesia side effects in Parkinson's patients with neural transplants. *Sci. Transl. Med.* **2**, 38–46 (2010).
3. Kim, J. H. *et al.* Dopamine neurons derived from embryonic stem cells function in an animal model of Parkinson's disease. *Nature* **418**, 50–56 (2002).
4. Barberi, T. *et al.* Neural subtype specification of fertilization and nuclear transfer embryonic stem cells and application in parkinsonian mice. *Nature Biotechnol.* **21**, 1200–1207 (2003).
5. Wernig, M. *et al.* Neurons derived from reprogrammed fibroblasts functionally integrate into the fetal brain and improve symptoms of rats with Parkinson's disease. *Proc. Natl Acad. Sci. USA* **105**, 5856–5861 (2008).

6. Heins, N. *et al.* Glial cells generate neurons: the role of the transcription factor Pax6. *Nature Neurosci.* **5**, 308–315 (2002).
7. Vierbuchen, T. *et al.* Direct conversion of fibroblasts to functional neurons by defined factors. *Nature* **463**, 1035–1041 (2010).
8. Ang, S.-L. Transcriptional control of midbrain dopaminergic neuron development. *Development* **133**, 3499–3506 (2006).
9. Smidt, M. P. & Burbach, J. P. How to make a mesodiencephalic dopaminergic neuron. *Nature Rev. Neurosci.* **8**, 21–32 (2007).
10. Sawamoto, K. *et al.* Visualization, direct isolation, and transplantation of midbrain dopaminergic neurons. *Proc. Natl Acad. Sci. USA* **98**, 6423–6428 (2001).
11. Perlmann, T. & Wallén-Mackenzie, A. Nurr1, an orphan nuclear receptor with essential functions in developing dopamine cells. *Cell Tissue Res.* **318**, 45–52 (2004).
12. Stadtfeld, M., Maherali, N., Breault, D. T. & Hochedlinger, K. Defining molecular cornerstones during fibroblast to iPS cell reprogramming in mouse. *Cell Stem Cell* **2**, 230–240 (2008).
13. Grace, A. A. & Bunney, B. S. The control of firing pattern in nigral dopamine neurons: single spike firing. *J. Neurosci.* **4**, 2866–2876 (1984).
14. Grace, A. A. & Onn, S. P. Morphology and electrophysiological properties of immunocytochemically identified rat dopamine neurons recorded *in vitro*. *J. Neurosci.* **9**, 3463–3481 (1989).
15. Pothos, E. N., Davila, V. & Sulzer, D. Presynaptic recording of quanta from midbrain dopamine neurons and modulation of the quantal size. *J. Neurosci.* **18**, 4106–4118 (1998).
16. Staal, R. G. W., Mosharov, E. V. & Sulzer, D. Dopamine neurons release transmitter via a flickering fusion pore. *Nature Neurosci.* **7**, 341–346 (2004).
17. Simeone, A. Genetic control of dopaminergic neuron differentiation. *Trends Neurosci.* **28**, 62–65 (2005).
18. Zappone, M. V. *et al.* Sox2 regulatory sequences direct expression of a β -geo transgene to telencephalic neural stem cells and precursors of the mouse embryo, revealing regionalization of gene expression in CNS stem cells. *Development* **127**, 2367–2382 (2000).
19. Pang, Z. P. *et al.* Induction of human neuronal cells by defined transcription factors. *Nature* advance online publication, doi:10.1038/nature10202 (26 May 2011).
20. Pfisterer, U. *et al.* Direct conversion of human fibroblasts to dopaminergic neurons. *Proc. Natl Acad. Sci. USA* doi:10.1073/pnas.1105135108 (6 June 2011).
21. Amarglio, N. *et al.* Donor-derived brain tumor following neural stem cell transplantation in an ataxia telangiectasia patient. *PLoS Med.* **6**, e1000029 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are thankful to D. Bonanomi, S.-L. Ang, S. El Mestikawy, M. P. Smidt, M. German and F. Valtorta for providing valuable antibodies. We thank A. Sessa and V.B. laboratory members for helpful discussion. M. Wernig is acknowledged for providing the iN-inducing lentiviral vectors. We are thankful to S. Nicolis for sharing Sox2^{β-geo} mice. L. Muzio, C. Laterza and G. Martino are acknowledged for the generation of Sox2^{β-geo} induced pluripotent stem cells. M. Bacigaluppi is acknowledged for advice on stereological countings. We thank the “Cell Line and DNA Biobank” (G. Gaslini Institute) and “Human Genetic Bank of Patients affected by Parkinson Disease and parkinsonism” (Parkinson Institute of Milan) of the Teletthon Genetic Biobank Network for human fibroblast samples. This study was supported by the “Fondazione Grigioni per il Morbo di Parkinson” (grant no. FGBRCVNI10310-001-V.B.), Eranet Neuron (V.B.), Cariplo Foundation (V.B.), Ministry of Health (Giovani ricercatori Award) (V.B.) and Italian Institute of Technology (V.B., A.D., S.G., T.S., R.G.).

Author Contributions M.C. and V.B. designed and conceived the experiments. M.C., M.T.D. and G.C. performed the lentiviral infections, characterized reprogrammed cells and analysed their fate after *in vivo* transplantation. E.D. and A.D. designed, performed and analysed all electrophysiological experiments. P.R., D.L., P.C. and S.G. performed the microarray gene expression profiling and analysed the data. D.L., A.D. and R.R.G. designed and D.L. and E.D. performed the amperometric experiments. R.R.G. and T.D.S. designed the protocol and performed the assessment of dopamine levels. S.T. and G.R. performed patch-clamp recording on brain slices. A.M. performed the functional analysis of synaptic activity. G.P. supervised the selection of the Parkinson's disease patients and the isolation of the primary fibroblasts. V.B. and A.D. should be considered as co-senior authors and wrote the manuscript.

Author Information Data have been deposited in NCBI's Gene Expression Omnibus and are accessible through GEO series accession number GSE27174 (<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE27174>). Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to V.B. (broccoli.vania@hsr.it).

METHODS

Cell culture. MEFs were isolated from E14.5 wild-type or TH-GFP knock-in mice embryos. Head, vertebral column, dorsal root ganglia and all internal organs were removed and discarded and the remaining embryonic tissue was manually dissociated and incubated in 0.25% trypsin (Sigma) for 10–15 min. Cells from each embryo were plated onto a 15-cm tissue culture dish in MEF media (Dulbecco's modified Eagle medium (DMEM; Invitrogen) containing 10% fetal bovine serum (FBS; Hyclone), non-essential amino acids (Invitrogen), sodium pyruvate and penicillin/streptomycin (Invitrogen)). In all experiments cells were not split more than four times. Mouse adult fibroblasts were isolated from tail tip samples. Tails were peeled, minced into 1 cm pieces, placed on culture dishes, and incubated in MEF media for 5 days. Adult human fibroblasts were isolated from skin biopsy samples of healthy and Parkinson's disease patients^{22,23} provided from the "Cell Line and DNA Biobank from Patients affected by Genetic Diseases" (G. Gaslini Institute) and "Parkinson Institute Biobank" (Milan, <http://www.parkinson.it/dnabank.html>) of the Telethon Genetic Biobank Network (<http://www.biobanknetwork.org>). The informed consent as issued by the ICP Ethical committee was obtained by healthy and Parkinson's disease patients enrolled for the DNA and cell biobank collection.

Human skin samples were mechanically dissociated and plated on matrigel-coated dishes. Human fibroblasts were cultured as MEFs. Mouse and adult fibroblasts were grown in MEF media as well as human fetal lung fibroblasts IMR90 (ATCC). Mesencephalic dopaminergic primary cell cultures from TH-GFP mice were prepared as previously described²⁴. Mice were maintained at San Raffaele Scientific Institute Institutional mouse facility and experiments were performed in accordance with experimental protocols approved by local Institutional Animal Care and Use Committees.

Molecular cloning and viral infection. Complementary DNAs for the dopaminergic transcription factors were cloned into lentiviral vectors under the control of the tetracycline operator⁷. Replication-incompetent, VSVg-coated lentiviral particles were packaged in 293T cells. MEFs, IMR90 and adult mouse and human fibroblasts were infected in MEF media. 16–20 h after infection cells were switched into fresh MEF media containing doxycycline (2 mg ml⁻¹; Sigma). After 48 h medium was replaced with neuronal inducing media (DMEM/F12 (Invitrogen), 25 µg ml⁻¹ insulin (Sigma), 50 µg ml⁻¹ transferrin (Sigma), 30 nM sodium selenite, 20 nM progesterone (Sigma), 100 nM putrescine (Sigma) and penicillin/streptomycin (Sigma)) containing doxycycline. The medium was changed every 2–3 days for a further 10–22 days. For proliferation assay, MEFs were treated with a 48 h pulse of 10 µM BrdU.

Immunohistochemistry. For immunocytochemical analysis, 5 × 10⁴ mouse or human fibroblasts were plated on matrigel-coated glass coverslips the day before the infection. 10–28 days after viral infection cells were fixed for 20 min at room temperature (20–25 °C) in 4% paraformaldehyde in PBS, permeabilized for 30 min in PBS containing 0.1% Triton X-100 and 10% normal goat serum (NGS), and incubated overnight at 4 °C in PBS containing 10% NGS and primary antibodies. Then cells were washed three times with PBS and incubated for 2 h at room temperature with anti-rabbit or anti-mouse secondary antibodies Alexa Fluor-488 or Alexa Fluor-594 (1:500; Invitrogen). For immunohistochemical analysis P15 or adult mouse brains were fixed overnight at 4 °C with 4% paraformaldehyde, buffered in 30% sucrose and embedded in OCT. Frozen brains were sectioned into 15- or 40-µm-thick sections with a cryostat and processed for immunostaining. Sections were boiled for 3 min in 10 mM citrate buffer solution pH 6 for antigen retrieval and permeabilized for 1 h at room temperature in PBS containing 0.1% or 0.25% Triton X-100 and 10% NGS. Primary antibodies were as follows: mouse anti-TH (1:200; Millipore), rabbit anti-TH (1:200; Immunological Sciences), mouse anti-βIII-tubulin (1:500; Covance), rabbit anti-βIII-tubulin (TuJ1) (1:500; Covance), rabbit anti-VMAT2 (1:200; Chemicon), rat anti-DAT (1:500; Millipore), rabbit anti-D2 receptor (1:100; Millipore), rabbit anti-calbindin (1:200; Swant), rabbit anti-AADC (1:100; Novus Biologicals), rabbit anti-ALDH1A1 (1:200; Abcam), mouse anti-synaptotagmin I (1:200; Synaptic Systems), mouse anti-synapsin (1:200; Synaptic Systems), chicken anti-GFP (1:2,000; Molecular Probes), rat anti-BrdU (1:200; BD), mouse anti-MAP2 (1:500; Immunological Sciences), rabbit anti-Otx2 (1:100 R&D). β-Galactosidase staining was performed as previously described²⁵.

Statistical analysis. The total numbers of Th⁺ and TuJ1⁺ cells were quantified 12–24 days after infection. Cell counting was performed on ten fields from three replicates for each condition and normalized with the number of cells plated before the infection. Data were expressed as mean ± s.e.m.

RT-PCR. RNA was extracted from single cultures, using Trizol isolation system (Invitrogen) according to manufacturer's instructions. The yield and integrity of the RNA were determined by the spectrophotometric measurement of A260 and by agarose-gel electrophoresis, respectively. Total RNA was treated with DNase I (Qiagen) to prevent DNA contamination. Two micrograms of RNA were reverse

transcribed using the Transcriptor High Fidelity cDNA Synthesis Kit (Roche). One twentieth of the reverse transcribed cDNA was amplified in a 25 microlitres of reaction mixture containing Taq polymerase buffer (Fisher BioReagents), 0.2 mM dNTPs (Finnzymes OY), 0.4 micromolar of each primer, 1 U Taq polymerase (Fisher BioReagents). The primers used to amplify cDNA samples are listed in Supplementary Table 6.

Cell sorting, laser capture dissection and microarray analysis. TH-GFP-positive iDA cells were directly sorted in Trizol (Invitrogen) using the cell sorter FACS Vantage s.e. DiVa (Becton Dickinson). Thus RNA was extracted as reported earlier and biotin-labelled cRNA was obtained using the Ovation kit (NuGEN). Labelled cRNA was hybridized (CBM genexpression facility, SISA) on Affymetrix Mouse Gene 1.0 ST Arrays, containing 35,557 probe sets corresponding to 28,853 genes. Hybridized arrays were stained and washed (GeneChip Fluidics Station 450) and scanned (GeneChip Scanner 3000 7G). Cell intensity values were computed using the Affymetrix GeneChip Operating Software (GCOS). Further data processing was performed in the R computing environment (<http://www.r-project.org/>) version 2.8.0 with BioConductor packages (<http://www.bioconductor.org/>). Robust Multi-Array Average (RMA) normalization was applied²⁶. Data were then filtered based on probe set intensity, so that only probe sets that had intensity values of >50 in at least half the arrays were retained. Statistical analysis was performed with limma²⁷. *P* values were adjusted for multiple testing using Benjamini and Hochberg's method to control the false discovery rate²⁸. Genes with adjusted *P* values below 0.01 were considered to be differentially expressed. Furthermore, a fold-change threshold cutoff was set to focus on genes whose expression level changed at least 2 times. Data were analysed through DAVID Bioinformatics Resources v6.7 (refs 28, 29).

Gene expression profiles of adult A9 and A10 dopaminergic neurons were obtained as previously described³⁰. In brief, adult TH-GFP female mice were killed by cervical dislocation. The brains were rapidly cut to isolate the midbrain region, and immediately immersed in 1× zinc fixative (BD Pharmingen) for 4–6 h at +4 °C. Once fixed, tissues were moved to 30% sucrose in 1× zinc fixative solution at +4 °C overnight. Following inclusion in OCT, tissues were frozen in iso-pentan (Sigma) and percolated with liquid nitrogen. 14-mm cryosections were mounted on SuperFrost plus glass slides (Menzel-Gläser) and air dried. mDA A9 and A10 neurons (each one from three different mice) were isolated from cryosections by using a PALM LCM microdissection system (PALM Microlaser Technology). To facilitate detection of fluorescent neurons, a drop of 1× zinc fixative was applied to the section during cell selection. The sections were air dried, neurons were dissected and catapulted onto PALM adhesive caps (Zeiss). Total RNA from 2,500 pooled neurons was isolated by using the Nano RNA extraction kit (Stratagene) and contaminating genomic DNA was removed through on-column DNase digestion step. The common reference RNA was generated from three midbrain regions of age-matched female mice. Midbrain RNA was isolated using RNeasy Mini kit (Qiagen), followed by DNase treatment. RNA from dissected neurons and all midbrains was amplified and labelled by Ovation Pico kit, WT exon and Encore biotin labelling kit (Nugene), following manufacturer's instructions. Once prepared, each target was hybridized on MoExon 1.0ST GeneChip (Affymetrix). Statistical analysis was performed by oneChannelGUI R package. All hierarchical clusters were generated by TMEV software.

Bisulphite genomic sequencing. DNA from sorted TH-GFP⁺ reprogrammed MEFs was modified using the CpGenome modification kit (Chemicon) according to the manufacturer's recommendations. Thus *Th* and *Vmat2* promoters CpG-rich selected regions were amplified using the PCR primers listed in Supplementary Table 5.

HPLC. To quantify dopamine level in reprogrammed cells, cell pellets were homogenized in 100 µl 0.1 N HClO₄ and analysed by using HPLC with electrochemical detection (Alexis 100, Antec Leyden). To measure dopamine concentrations in the supernatants, cells were exposed to media with or without 50 mM KCl for 30 min, then 0.9 ml of supernatants were collected with the addition of 0.1 ml of 1 N HClO₄, filtered and analysed by HPLC. Dopamine was separated on a reverse-phase column (ALB-105, 3 µm, 50 × 1 mm) with a mobile phase consisting of 50 mM phosphate buffer, 8 mM KCl, 500 mg l⁻¹ octyl sodium sulphate, 0.1 mM EDTA, and 3% methanol (pH 6.0) at a flow rate of 50 µl min⁻¹. Dopamine was detected by a Decade II electrochemical detector equipped with micro VT-03 electrochemical flow cell and a 0.7-mm-diameter glassy carbon electrode (Alexis 100, Antec Leyden). The volume of the injection was 5 µl. The detection limit established as a 3:1 signal-to-noise ratio was below 0.5 nM.

Electrophysiology. Recordings were performed from reprogrammed mouse and human fibroblasts and primary mDA neurons. The mouse TH-GFP⁺ cells selected for the electrophysiological analysis were not so flat as the fibroblasts, and had several well-developed neurites. The human cells selected for the electrophysiological analysis also had neuron-like shapes with clearly distinguishable neurites by phase contrast microscopy. Only cells without signs of detachment

from the substrate were used for recordings. Cells were perfused continuously with HEPES-buffered saline (HBS) of the following composition (in mM): 140 NaCl, 5 KCl, 2 CaCl₂, 2 MgCl₂, 15 HEPES, and 25 glucose, pH 7.4. The patch pipette solution contained (in mM): 130 K-gluconate, 10 KCl, 0.5 CaCl₂, 15 HEPES, 5 EGTA, 8 NaCl, 2 MgATP, 0.3 Na₂GTP, and 10 glucose, pH adjusted to 7.2 with KOH. Action potentials were recorded in the on-cell and the current-clamp whole-cell configurations. A current was injected to have membrane potentials around -60 mV, and step currents from -50 pA to 40 pA were injected to elicit action potentials. Na⁺ currents and composite K⁺ currents were recorded in the voltage-clamp configuration by delivering voltage steps ranging from -100 mV to +20 mV in cells held at -60 mV. Delayed rectifier K⁺ currents were activated by 0.5 s voltage steps from -40 mV to +20 mV after a 0.5-s-long step to -40 mV. A-type K⁺ currents were isolated by subtraction of delayed rectifier K⁺ currents from those activated by voltage steps after a 0.5-s-long step to -100 mV. Recordings were performed using an EPC10 USB patch clamp amplifier and PATCHMASTER software (HEKA Elektronik). Data were digitized at 10 kHz and analysed with FITMASTER Software (HEKA Elektronik). Detection and measurements of action potentials were performed using MiniAnalysis software (Synaptosoft).

In vivo electrophysiology. Slices were obtained from transplanted mice at postnatal day 42. The brains were quickly removed from the skull in ice-cold artificial cerebrospinal fluid (ACSF) containing the following (in mM): 125 NaCl, 25 NaHCO₃, 2.5 KCl, 1.25 NaH₂PO₄, 2 CaCl₂, 1 MgCl₂, and 25 glucose, pH 7.4 (bubbled with 95% O₂ and 5% CO₂). Coronal slices (300-μm thick) were cut using a vibratome (VT1000S; Leica) and stored in ACSF at 25–28 °C. For recording, slices were transferred to a recording chamber continuously superfused with ACSF (1–2 ml min⁻¹ at 30–32 °C). Whole-cell recordings were performed in both current- and voltage-clamp configurations. Recording pipettes (3–5 MΩ of resistance) contained the following solution (in mM): 124 KH₂PO₄, 10 NaCl, 2 MgCl₂, 0.5 EGTA, 10 HEPES, 2 Na₂-ATP, 0.03 Na-GTP (pH 7.2, adjusted with KOH). Signals were sampled at 10 kHz, filtered at 2 kHz, and acquired using a MultiClamp 700A amplifier and pClamp 10 software (Molecular Devices).

Amperometric recording. Amperometry was used to detect the evoked catecholamine exocytosis from single cells^{15,31}. Carbon-fibre microelectrodes were fabricated from 5 μm carbon fibres (Goodfellow), inserted in a 1.2 × 0.68 mm glass capillary (A-M system, Sequim) and pulled with a PE-22 micropipette puller (Narishige). Electrodes were sealed by dipping in Epoxy resin (Epo-Tek 301, Epoxy Technology) and cured at 100 °C for 24 h. They were backfilled with 3 M KCl and trimmed to obtain a basal current between 140 and 180 nA. The electrodes' responses were tested by cyclic voltammetry and those with unstable cyclic voltammograms, when tested in a solution of 10 μM dopamine, were rejected. A voltage was applied to the carbon fibre using an EPC10 USB patch-clamp amplifier (HEKA Elektronik). The signal was low-pass filtered at 10 kHz using a 4-pole Bessel filter, digitalized at 50 kHz and digitally refiltered at 1–1,000 Hz. The latter resulted in slightly longer responses, but significantly improved the visualization of secretion events that was the only aim of these experiments. The electrode was positioned adjacent to individual cells and lowered to approach the somatodendritic domains of iDA cells³², using an Olympus BX51WI microscope with ×40 water immersion objective. To increase the signal-to-noise ratio, cells were pre-treated with 100 μM L-DOPA (Sigma-Aldrich) for 30 min although we were able to resolve single spike-like release events in two untreated cells. The experiments consisted of current recordings at +750 mV during a brief baseline period, during which cells were perfused with standard external medium containing 5 mM K⁺. It was then exchanged for a stimulation solution (25 mM K⁺), and amperometric signals were recorded for a further period of 7 min. Catecholamine secretion was apparent as discrete spike-like events, each corresponding to vesicular catecholamine release. Most events were detected during 2 min after 25 mM K⁺ stimulation, but occasional events were observed also during baseline recordings. No vesicular release of dopamine was recorded at the electrode placed adjacent to a cell when the applied potential was 0 mV or -750 mV, or at +750 mV when the electrode was placed remotely from cells.

FM4-64 assay. FM4-64 dye uptake experiments were performed as previously reported³³. Briefly, 21 days *in vitro* TH-GFP⁺ iDA cells were stimulated for 1 min with 55 mM KCl, in the presence of FM4-64 (10 μM). After FM4-64 loading, neuronal cells were washed and perfused for 10 min with warmed Krebs buffer (37 °C) supplemented with TTX (1 μM) and CNQX (10 μM). After live fluorescent FM4-64 signals were acquired, cells were fixed and immunostained for TH and SYT1.

Electron microscopy. For ultrastructural immunocytochemistry, 21 days *in vitro* infected MEFs were fixed in 2% glutaraldehyde in PBS, washed in PBS, postfixed in 2% OsO₄ in PBS, and embedded in Epon. Ultrathin sections prepared from these samples were analysed with electron microscope (H-7000; Hitachi).

Transplantation surgery. After 4 days of infection, TH-GFP MEFs were trypsinized and resuspended at 2×10^5 cells μl⁻¹ in fresh prepared Krebs buffer containing the following (in mM): 126 NaCl, 2.5 KCl, 1.2 NaH₂PO₄, 1.2 MgCl₂, 2.1 CaCl₂, 11 glucose, 4.2 NaHCO₃, 1 HEPES, and 1% vital dye Fast Green. P1 mice pups were anaesthetized by hypothermia (4 min) and fixed to a support using band-aid. The skin and the skull overlying the lateral ventricle were opened over about 2 mm using an ophthalmic scalpel. Subsequently, the mouse was placed in a stereotaxic rig (Kopff) under a Hamilton syringe containing 2 μl of cell suspension. The syringe was placed over the incision, positioned at the level of the skull, then lowered into the lumen of the right lateral ventricle (LV; 2.5 mm) or in the somatosensory cortex (1.5 mm) and cell solution was injected. Mice were left on a 37 °C heating blanket for several minutes after surgical manipulation to avoid fatal hypothermia.

Stereological analysis. Three mice transplanted with reprogrammed cells were used for stereological analysis. Three weeks after transplantation mice were anaesthetized and killed by transcardiac perfusion with PBS followed by 4% paraformaldehyde. Brains were cryoprotected through incubation in an ice-cold solution of 30% sucrose in PBS and cut in coronal 40-μm-thick cryostat sections. From these sections, one systematic random series of sections was stained for GFP, so that sections were spaced at 7 section intervals (total of 16 sections per mouse). GFP immunoperoxidase staining was performed as described elsewhere³⁴. Cells were quantified using the assistance of the Stereo Investigator v 3.0 software (MicroBrightField) and a personal computer running the software connected to a colour video camera mounted on a Leica microscope^{35,36}. The motorized stage of the microscope allowed precise and well-defined movements along the x-, y- and z-axes. Images were first acquired with a CCD-IRIS colour video camera and the cerebral hemispheres were interactively delineated at low magnification on a video image of the section. Counting of cells was performed manually on every seventh section using a ×40 lens. To estimate the total number of GFP positive cells the total number of neurons counted on the sections was multiplied by seven.

22. Sironi, F. *et al.* α-Synuclein multiplication analysis in Italian familial Parkinson disease. *Parkinsonism Relat. Disord.* **16**, 228–231 (2010).
23. Sironi, F. *et al.* Parkin analysis in early onset Parkinson's disease. *Parkinsonism Relat. Disord.* **14**, 326–333 (2008).
24. Pruszek, J. *et al.* Isolation and culture of ventral mesencephalic precursor cells and dopaminergic neurons from rodent brain. *Curr. Prot. Stem Cell Biol.* Chapter 2, 2D.5.1–2D.5.21 (2009).
25. Broccoli, V. *et al.* The caudal limit of *Otx2* expression positions the isthmus organizer. *Nature* **401**, 164–168 (1999).
26. Irizarry, R. A. *et al.* Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res.* **31**, e15 (2003).
27. Smyth, G. K. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* **3**, Article 3 (2004).
28. Hochberg, Y. & Benjamini, Y. More powerful procedures for multiple significance testing. *Stat. Med.* **9**, 811–818 (1990).
29. Huang, D. W. *et al.* Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37**, 1–13 (2009).
30. Biagioli, M. *et al.* Unexpected expression of α- and β-globin in mesencephalic dopaminergic neurons and glial cells. *Proc. Natl Acad. USA* **106**, 15454–15459 (2009).
31. Pothos, E., Desmond, D. & Sulzer, D. L-3,4-Dihydroxyphenylalanine increases the quantal size of exocytotic dopamine release *in vitro*. *J. Neurochem.* **66**, 629–636 (1996).
32. Mundroff, M. L. & Wightman, R. M. Amperometry and cyclic voltammetry with carbon fiber microelectrodes at single cells. *Curr. Protoc. Neurosci.* Chapter 6, 6.14.1–6.14.22 (2002).
33. Menegon, A. *et al.* Protein kinase A-mediated synapsin I phosphorylation is a central modulator of Ca²⁺-dependent synaptic activity. *J. Neurosci.* **26**, 11670–11681 (2006).
34. Colasante, G. *et al.* Arx is a direct target of Dlx2 and thereby contributes to the tangential migration of GABAergic interneurons. *J. Neurosci.* **28**, 10674–10686 (2008).
35. West, M. J. *et al.* Unbiased stereological estimation of the total number of neurons in the subdivisions of the rat hippocampus using the optical fractionator. *Anat. Rec.* **231**, 482–497 (1991).
36. Bacigaluppi, M. *et al.* Delayed post-ischaemic neuroprotection following systemic neural stem cell transplantation involves multiple mechanisms. *Brain* **132**, 2239–2251 (2009).

Legionella pneumophila SidD is a deAMPyase that modifies Rab1

Yunhao Tan¹ & Zhao-Qing Luo¹

Legionella pneumophila actively modulates host vesicle trafficking pathways to facilitate its intracellular replication with effectors translocated by the Dot/Icm type IV secretion system (T4SS)¹. The SidM/DrrA protein functions by locking the small GTPase Rab1 into an active form by its guanine nucleotide exchange factor (GEF) and AMPylation activity^{2–4}. Here we demonstrate that the *L. pneumophila* protein SidD preferably deAMPyates Rab1. We found that the deAMPylation activity of SidD could suppress the toxicity of SidM to yeast and is required to release Rab1 from bacterial phagosomes efficiently. A molecular mechanism for the temporal control of Rab1 activity in different phases of *L. pneumophila* infection is thus established. These observations indicate that AMPylation-mediated signal transduction is a reversible process regulated by specific enzymes.

L. pneumophila, the aetiological agent of Legionnaires' disease, is capable of colonizing a wide range of eukaryotic cells. Successful infection by this pathogen requires the Dot/Icm T4SS, which translocates numerous protein substrates into host cells¹. These proteins modulate various host cellular pathways, such as vesicle trafficking, cell death, lipid metabolism and protein synthesis, to construct a phagosome permissive for intracellular bacterial replication⁵. Accumulating evidence suggests that a unique repertoire of effectors is required for each phase of the infection. There is a need for the bacterium to regulate the activity of its virulence factor, because some effectors are potentially detrimental to host cells. Such regulation can be achieved by various mechanisms, including the control of the amount of protein delivered into host cells, the stability of the protein or susceptibility of the protein to host degradation machinery. *L. pneumophila* has evolved unique mechanisms to neutralize the activity of effectors whose functions presumably have become detrimental to, or no longer important for, the development of the bacterial phagosome. For example, several hours after infection, the bacterial E3 ubiquitin ligase, LubX, targets the effector SidH for proteasomal degradation⁶. Similarly, several hours after uptake, LepB, a GTPase activation protein (GAP) for Rab1, antagonizes the effects of SidM (also known as DrrA), which recruits the small GTPase to the *Legionella*-containing vacuole (LCV) and converts it into the active GTP-bound form via its guanine nucleotide exchange factor (GEF) activity^{2,3,7}. SidM also catalyses an adenosine monophosphate modification (AMPylation) on the tyrosine 77 of Rab1 to lock it into the active form⁴. Post-translational modification by AMPylation has recently emerged as a novel cellular signalling mechanism used by all domains of organisms^{8,9}. However, little is known about the regulation of this signalling mechanism and naturally occurring enzymes involved in the reversal of the modification remain elusive.

In an earlier study we isolated a number of Dot/Icm substrates toxic to yeast, such as SidI, Lgts, SidM and AnkX^{10–12}. To determine whether the activity of any of these proteins is under direct regulation of bacterial factors, we initiated screenings to identify *L. pneumophila* proteins capable of suppressing the toxicity to yeast. A plasmid-borne *L. pneumophila* genomic library was introduced into yeast strains expressing toxic effectors from a galactose-inducible promoter, leading to the identification of a number of clones that efficiently suppress the

toxicity of SidM. Sequencing revealed that all of these clones harboured *sidD* (*lpg2465*), which encodes a Dot/Icm substrate of 507 amino acids¹³. In the *L. pneumophila* genome, *sidD* is localized next to *sidM* and these two genes are transcribed in divergent orientations^{2,13,14}.

Co-expression of *sidD* completely rescued the growth of the SidM-producing yeast strain on inducing media (Fig. 1a). SidD was unable to suppress the toxicity of AnkX, which is believed to interfere with host vesicle trafficking by AMPylating substrate(s) as yet unidentified in a Fic-domain-dependent manner¹⁵ (Supplementary Fig. 1), indicating that the suppressor activity of SidD is specific for SidM.

SidM is a protein of multiple functions, which by binding to phosphatidylinositol 4'-monophosphate, anchors on the *Legionella* vacuole, recruits and activates the small GTPase Rab1^{2,3,16}. In particular, its amino-terminal domain (SidM1–339) possesses an adenosine monophosphorylation (AMPylation) activity, which covalently modifies Rab1 at tyrosine 77 in a process that requires the G₉₈X₍₁₁₎D₁₁₀XD₁₁₂ motif conserved between SidM and the glutamine synthetase adenylyl transferase⁴. We thus examined whether SidD is able to suppress the toxicity induced by SidM1–339. Expression of SidM1–339 strongly inhibited yeast growth (Fig. 1a, strain F) and such inhibition can be suppressed by SidD (Fig. 1a, strain G), indicating that SidD interferes with the activity conferred by the AMPylation function. We also examined the ability of SidD to suppress the AMPylation-dependent toxicity of SidM1–339 in mammalian 293T cells⁴. Transfection of these cells with green fluorescent protein-conjugated SidM1–339 (GFP-SidM1–339) led to extensive cell rounding. Cotransfection of the cells to express Flag-SidD suppressed this toxicity (Fig. 1c–e). When five times more DNA of the SidD-expressing plasmid was used, at least 85% of the cells were protected (Fig. 1c–e). Together, these data suggest that SidD is able to neutralize the effect of SidM in eukaryotic cells either by inhibiting its AMPylation activity or by reversing its effects on the target molecule Rab1.

To determine the mechanism of action of SidD, we tested its effect on the AMPylation activity of SidM with purified recombinant proteins. Incubation of glutathione-S-transferase-conjugated SidM (GST-SidM) with GST-Rab1 in the presence of [α -³²P]ATP for 30 min led to robust production of radiolabelled Rab1 (Fig. 2a, lane 1). Consistent with the genetic data, inclusion of His₆-SidD in the AMPylation reaction abolished the formation of [α -³²P]AMP-Rab1 (Fig. 2a, lane 2). The lack of AMPylation can be a result of SidD-mediated inhibition of SidM activity or of a deAMPylation function of SidD. Because full-length SidM has a high binding affinity for Rab1 (refs 2–4), its presence may interfere with SidD activity. We distinguished between these two possibilities by producing [α -³²P]AMP-Rab1 using His₆-SidM1–339, which catalyses the reaction but did not detectably bind GST-Rab1 *in vitro*^{2–4} (Supplementary Fig. 2). After removing His₆-SidM1–339, AMPylated Rab1 was incubated with various amounts of His₆-SidD for 5 min. Under these conditions, the amount of radiolabelled GST-Rab1 significantly decreased when as little as 0.6 μ g His₆-SidD was included in the reaction, and 1.2 μ g His₆-SidD can almost completely deAMPyate [α -³²P]AMP-Rab1 (Fig. 2b). In a time course analysis, incubation of 1.0 μ g SidD with 10 μ g modified Rab1 for 1 min led to more than a 50%

¹Department of Biological Sciences, Purdue University, 915 West State Street, West Lafayette, Indiana 47907, USA.

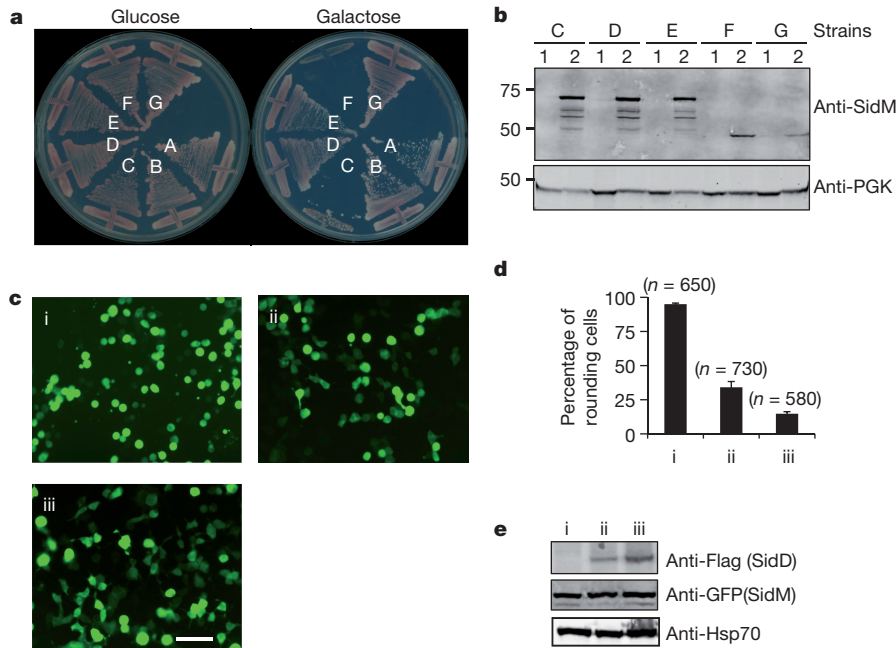


Figure 1 | Suppression of the cytotoxicity of SidM by SidD. **a**, Suppression of yeast toxicity of SidM. Yeast strains expressing SidM or SidM1–339 from a galactose-inducible promoter were transformed with various plasmids harbouring *sidD* and the cells were streaked onto plates containing glucose or galactose. Plates were incubated at 30 °C for 3 days before acquiring the images. Yeast strains: A, vector/vector; B, vector/pSidD; C, pSidM/vector; D, pSidM/pSidD (original clone number 1); E, pSidM/pSidD; F, pSidM1–339/vector; and G, pSidM1–339/pSidD. **b**, SidD did not affect the protein levels of SidM or SidM1–339 in yeast cells. Subcultures of relevant yeast strains were grown in raffinose (1) or in galactose (2) medium. Crude lysates resolved by SDS–polyacrylamide gel electrophoresis

(PAGE) were probed with SidM-specific antibody. The 3-phosphoglycerate kinase (PGK) was used as a loading control (bottom). Markers for protein size (kDa) are shown. **c–e**, Co-expression of SidD rescued the cell-rounding phenotypes caused by SidM1–339. 293T cells were transfected to express SidM1–339. A SidD plasmid was not included (i) or was used at 1:3 (ii), or 1:5 (iii) molar ratio, respectively. After transfection (24 h), samples were analysed by acquiring images (**c**), by enumerating green cells exhibiting the rounding phenotype (**d**) or by immunoblotting to examine the protein levels of SidM1–339 and SidD (**e**). Hsp70 was probed as a loading control. Experiments were repeated three times and similar results were obtained. Error bars indicate s.d. Scale bar, 50 μ m.

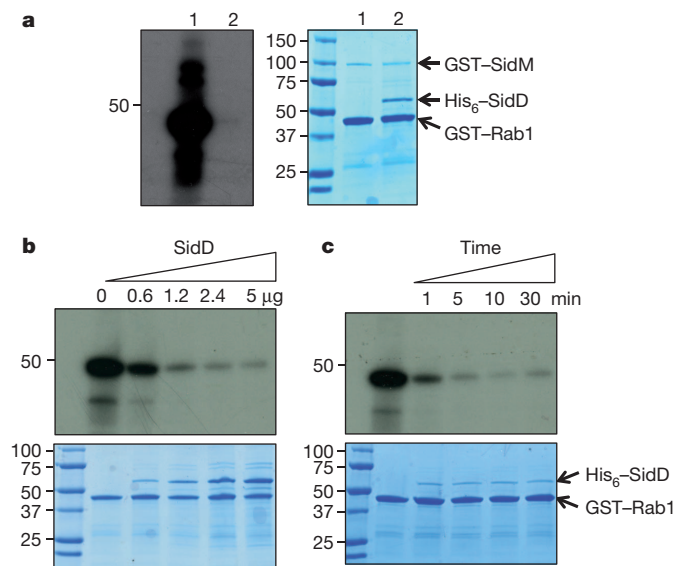


Figure 2 | SidD is a deAMPyase that targets SidM-modified Rab1. **a**, SidD prevented SidM-mediated AMPylation of Rab1. AMPylation reactions containing GST–SidM and GST–Rab1 without (lane 1) or with (lane 2) His₆–SidD are shown. After SDS–PAGE, [α -³²P]AMP–Rab1 was detected by autoradiography (left panel) and proteins were detected by Coomassie bright blue staining (right panel). **b**, Dose-dependent deAMPylation by SidD. His₆–SidD was added to identical samples containing AMPylated GST–Rab1 to establish reactions in which the molar ratio of Rab1 and SidD is 8, 4, 2 and 1, respectively; reactions were terminated after 5 min of incubation. AMPylated GST–Rab1 and proteins in reactions were detected as described in **a**. **c**, Time course of SidD activity. AMPylated GST–Rab1 was mixed with His₆–SidD at a molar ratio of 10:1, reactions were terminated at the indicated time intervals. AMPylated GST–Rab1 (upper panel) and proteins (bottom panel) were similarly detected. Markers for protein size (kDa) are shown.

reduction in radiolabelled substrate (Fig. 2c). SidD is also able to remove the AMP moiety from other Rab proteins modified by SidM⁴ (Supplementary Fig. 3).

To verify the SidD-mediated removal of the AMP moiety from modified Rab1, AMPylated Rab1 treated with His₆–SidD completely digested with trypsin was subjected to mass spectrometric analysis. The relative intensity ratio data showed that approximately 50% of the Rab1 was AMPylated when 10 μ g Rab1 was incubated with 1.5 μ g SidM for 30 min (Supplementary Fig. 4a). The ratio of unmodified Rab1 increased about tenfold when 4 μ g His₆–SidD was added to the reaction for 10 min and further increased to 20-fold when the incubation was extended for 30 min (Supplementary Fig. 4b, c). Furthermore, we did not detect any mass loss of the Rab1 peptide containing tyrosine 77, indicating that deAMPylation by SidD was not caused by other hydrolytic reaction. Taken together, these results indicate that SidD is a deAMPyase that functions to reverse the SidM-mediated post-translational modification of Rab1.

Sequence analysis with HHPred¹⁷ revealed that the predicted secondary structure of SidD N-terminal portion is detectably similar to some protein phosphatases (Supplementary Fig. 5). In particular, the Asp residues at positions 92 and 110 of SidD are conserved with residues critical for the biochemical activity of SaSTP and tPphA, phosphatases from *Streptococcus agalactiae* and *Thermosynechococcus elongates*, respectively^{18,19} (Supplementary Fig. 5). Substitution of D92 or D110 with alanine completely abolished the ability of SidD to suppress the toxicity of SidM to yeast, whereas a mutation in D60 did not detectably affect its activity (Fig. 3a, b). Consistently, SidD_{D92A} and SidD_{D110A} both had lost the activity to remove the AMP moiety from AMPylated Rab1 (Fig. 3c). These results indicate that residues D92 and D110 participate in the formation of the catalytic pocket or structural features important for its enzymatic function.

In bone-marrow-derived mouse macrophages, the association of Rab1 with LCVs peaks at 2 h after bacterial uptake and begins to

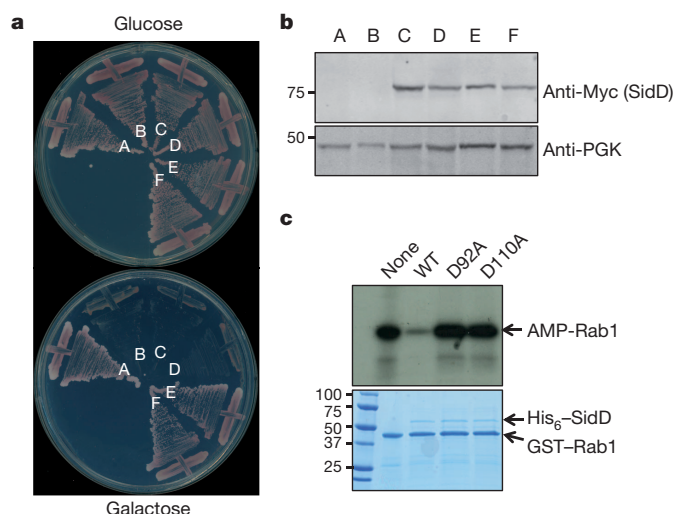


Figure 3 | The Asp residue at position 92 or 110 is important for SidD activity. **a**, Mutations of D92 and D110 abolished the ability of SidD to suppress the yeast cytotoxicity of SidM. Plasmids harbouring *sidD* or its mutants were transformed into the SidM-expression yeast strain; cells were streaked onto plates containing glucose or galactose. Yeast strains: A, vector/vector; B, pSidM/vector; C, pSidM/pSidD_{D92A}; D, pSidM/pSidD_{D110A}; E, pSidM/pSidD_{D60A} and F, pSidM/pSidD. **b**, Expression of the SidD mutants in yeast, samples were prepared as described in Fig. 1 and were probed for SidD (Myc-tagged) and for PGK. **c**, SidD_{D92A} and SidD_{D110A} have lost the deAMPylation activity. Purified proteins (1.5 µg) were added to reactions containing AMPylated Rab1. After 30 min incubation, reactions were terminated by SDS sample buffer. [α -³²P]AMP-GST-Rab1 was detected by autoradiography and the proteins were detected by Coomassie bright blue staining (lower panel). Protein size (kDa) references are shown on the left lane of the gel.

disappear when infection has proceeded for 4 h (ref. 7). The removal of Rab1 from LCVs is presumably due to extraction by RabGDI, which only interacts with the inactive GDP-bound form²⁰. LepB is unable to induce GTP hydrolysis of AMPylated Rab1, suggesting that a deAMPylation factor is necessary for the production of the GDP-bound, inactive Rab1 (ref. 4). Thus, we examined whether SidD is required for efficient removal of Rab1 from the LCVs. In infections with wild-type bacteria, the percentages of vacuoles staining positively for Rab1 began to diminish at 2 h (Fig. 4a). On the other hand, in infections with the *sidD* deletion mutant, the rates of Rab1-positive vacuoles remained at the maximal level (about 60%) 4 h after infection (Fig. 4). The kinetics of Rab1-positive LCVs was restored to the pattern of wild-type bacteria when SidD was expressed in the mutant (Fig. 4a, b). Interestingly, the association of Rab1 with the LCV did not persist after 4 h of infection with the *sidD* mutant (Fig. 4a), suggesting the contribution of host deAMPylation or additional bacterial proteins for reversal of the modification. These results indicate that SidD plays an important role in the efficient removal of Rab1 from the *Legionella* phagosome.

AMPylation of proteins often leads to alteration of their enzymatic activity or the ability to interact with target molecules^{4,8,21,22}. Proteins with AMPylation activity dependent upon the Fic domain are present in all kingdoms and at least one has been characterized in humans^{8,9}. Although the importance of signalling pathways regulated by AMPylation has not yet been fully recognized, the fact that this reaction can be catalysed by proteins lacking a Fic domain such as SidM suggests extensive involvement of this post-translational modification in cellular signalling processes. The identification of a deAMPylation revealed that like other post-translational modifications involved in signal transduction, AMPylation is a reversible process regulated by specific enzymes. Similar to *sidM*^{2,3} and *lepB*⁷, *sidD* is not required for bacterial intracellular growth¹³, probably because of functional redundancy among the several hundred Dot/Icm substrates²³ or

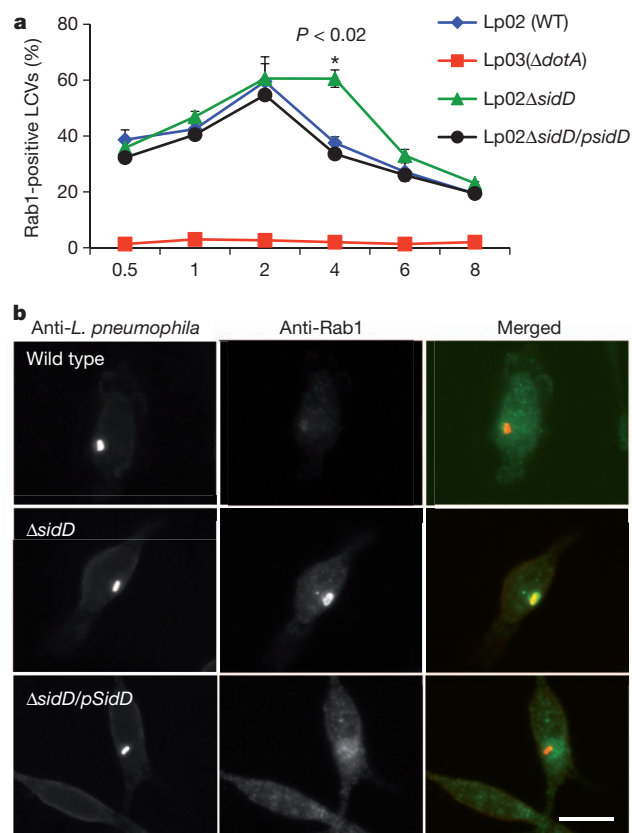


Figure 4 | SidD is required for efficient removal of Rab1 from *L. pneumophila* phagosome. **a**, Mouse macrophages were infected with relevant *L. pneumophila* strains. At the indicated time points, fixed samples were probed for *L. pneumophila* and Rab1 with specific antibodies followed by Texas red and FITC-conjugated secondary antibodies, respectively. Processed samples were scored for colocalization of Rab1 with the bacterial phagosomes. Data shown are from two independent experiments performed in triplicate in which at least 100 phagosomes were scored per sample. **b**, Association of Rab1 with *L. pneumophila* phagosome 4 h after infection. Shown are images of wild type (Lp02, *dot/icm*⁺), the *sidD* deletion mutant (Lp02 Δ sidD, *dot/icm*⁺) and the complementation strain (Lp02 Δ sidD/Flag-SidD) residing in macrophages 4 h after infection. *L. pneumophila* and Rab1 are labelled as described in **a**, with bacteria marked in red and Rab1 marked in green. Scale bar, 10 µm. Note the difference in the intensity of Rab1 staining signals among the three bacterial strains. At least 150 vacuoles were scored in each sample and error bars indicate s.d. Similar results were obtained in at least three independent experiments.

because of the limitation of the experimental systems. Nevertheless, further study on the structure and function of SidD will advance not only our understanding of its roles in bacterial pathogenesis but also the involvement of such enzymes in other cellular processes.

METHODS SUMMARY

Bacterial, yeast strains and plasmid construction. All *L. pneumophila* strains used in this study were derivatives of the Philadelphia 1 strain Lp02 (ref. 24). *Escherichia coli* strains were grown and maintained on LB agar or LB broth. When necessary antibiotics were included as described²⁵. Strains of *L. pneumophila* were grown and maintained on CYE medium or in AYE broth as previously described²⁵. The *sidD* in-frame deletion mutant was constructed in an earlier study¹³. In this mutant, the open reading frame of *sidD* was replaced by a 32-amino-acid polypeptide consisting of the first and the last 15 amino acids and two amino acids encoded by the sequence of the BamHI restriction enzyme¹³. For complementation experiments, a Flag-tagged *sidD* gene was inserted into the *sidD* locus of the deletion mutant using a two-step site-specific recombination with the π -protein-dependent plasmid pSR47s²⁴ with an established procedure²⁶. Successful insertion of the gene into the bacterial chromosome was determined by PCR reactions with Flag-tag-specific primers and by the expression of Flag-tagged SidD. All infections were performed with bacterial cultures grown to the post-exponential phase as judged by attenuation of the cultures (D_{600} = 3.3–3.8) as well

as increase of bacterial motility. For expression in mammalian cells, genes were cloned into pEGFPC1 (Clontech) or a 4×Flag vector (Sigma). The integrity of all constructs was verified by sequencing analysis. The sequences of all primers used in this study are listed in Supplementary Table 1.

All yeast strains used were derived from W303 (ref. 27); yeast was grown at 30 °C in YPD medium or in appropriate amino acid dropout synthetic media with glucose or galactose at a final concentration of 2% as the sole carbon source. Yeast transformation was performed according to a standard procedure²⁸.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 8 May; accepted 17 June 2011.

Published online 6 July 2011.

1. Ensminger, A. W. & Isberg, R. R. *Legionella pneumophila* Dot/Icm translocated substrates: a sum of parts. *Curr. Opin. Microbiol.* **12**, 67–73 (2009).
2. Machner, M. P. & Isberg, R. R. Targeting of host Rab GTPase function by the intravacuolar pathogen *Legionella pneumophila*. *Dev. Cell* **11**, 47–56 (2006).
3. Murata, T. *et al.* The *Legionella pneumophila* effector protein DrrA is a Rab1 guanine nucleotide-exchange factor. *Nature Cell Biol.* **8**, 971–977 (2006).
4. Muller, M. P. *et al.* The *Legionella* effector protein DrrA AMPylates the membrane traffic regulator Rab1b. *Science* **329**, 946–949 (2010).
5. Isberg, R. R., O'Connor, T. J. & Heidtman, M. The *Legionella pneumophila* replication vacuole: making a cosy niche inside host cells. *Nature Rev. Microbiol.* **7**, 13–24 (2009).
6. Kubori, T., Shinzawa, N., Kanuka, H. & Nagai, H. *Legionella* metaeffector exploits host proteasome to temporally regulate cognate effector. *PLoS Pathog.* **6**, e1001216 (2010).
7. Ingmundson, A., Delprato, A., Lambright, D. G. & Roy, C. R. *Legionella pneumophila* proteins that regulate Rab1 membrane cycling. *Nature* **450**, 365–369 (2007).
8. Worby, C. A. *et al.* The fic domain: regulation of cell signaling by adenylation. *Mol. Cell* **34**, 93–103 (2009).
9. Kinch, L. N., Yarbrough, M. L., Orth, K. & Grishin, N. V. Fido, a novel AMPylation domain common to Fic, Doc, and AvrB. *PLoS ONE* **4**, e5818 (2009).
10. Pan, X., Luhmann, A., Satoh, A., Laskowski-Arce, M. A. & Roy, C. R. Ankyrin repeat proteins comprise a diverse family of bacterial type IV effectors. *Science* **320**, 1651–1654 (2008).
11. Belyi, Y., Tabakova, I., Stahl, M. & Aktories, K. Lgt: a family of cytotoxic glucosyltransferases produced by *Legionella pneumophila*. *J. Bacteriol.* **190**, 3026–3035 (2008).
12. Shen, X. *et al.* Targeting eEF1A by a *Legionella pneumophila* effector leads to inhibition of protein synthesis and induction of host stress response. *Cell. Microbiol.* **11**, 911–926 (2009).
13. Luo, Z. Q. & Isberg, R. R. Multiple substrates of the *Legionella pneumophila* Dot/Icm system identified by interbacterial protein transfer. *Proc. Natl Acad. Sci. USA* **101**, 841–846 (2004).
14. Chien, M. *et al.* The genomic sequence of the accidental pathogen *Legionella pneumophila*. *Science* **305**, 1966–1968 (2004).
15. Roy, C. R. & Mukherjee, S. Bacterial FIC proteins AMP up infection. *Sci. Signal.* **2**, pe14 (2009).
16. Brombacher, E. *et al.* Rab1 guanine nucleotide exchange factor SidM is a major phosphatidylinositol 4-phosphate-binding effector protein of *Legionella pneumophila*. *J. Biol. Chem.* **284**, 4846–4856 (2009).
17. Soding, J. Protein homology detection by HMM–HMM comparison. *Bioinformatics* **21**, 951–960 (2005).
18. Rantanen, M. K., Lehtio, L., Rajagopal, L., Rubens, C. E. & Goldman, A. Structure of *Streptococcus agalactiae* serine/threonine phosphatase. The subdomain conformation is coupled to the binding of a third metal ion. *FEBS J.* **274**, 3128–3137 (2007).
19. Schlicker, C. *et al.* Structural analysis of the PP2C phosphatase tPpA from *Thermosynechococcus elongatus*: a flexible flap subdomain controls access to the catalytic site. *J. Mol. Biol.* **376**, 570–581 (2008).
20. Seabra, M. C. & Wasmeier, C. Controlling the location and activation of Rab GTPases. *Curr. Opin. Cell Biol.* **16**, 451–457 (2004).
21. Woolery, A. R., Luong, P., Broberg, C. A. & Orth, K. AMPylation: something old is new again. *Front. Microbiol.* **1**, 113 (2010).
22. Yarbrough, M. L. *et al.* AMPylation of Rho GTPases by *Vibrio* VopS disrupts effector binding and downstream signaling. *Science* **323**, 269–272 (2009).
23. Zhu, W. *et al.* Comprehensive identification of protein substrates of the Dot/Icm type IV transporter of *Legionella pneumophila*. *PLoS ONE* **6**, e17638 (2011).
24. Berger, K. H. & Isberg, R. R. Two distinct defects in intracellular growth complemented by a single genetic locus in *Legionella pneumophila*. *Mol. Microbiol.* **7**, 7–19 (1993).
25. Conover, G. M., Derre, I., Vogel, J. P. & Isberg, R. R. The *Legionella pneumophila* LidA protein: a translocated substrate of the Dot/Icm system associated with maintenance of bacterial integrity. *Mol. Microbiol.* **48**, 305–321 (2003).
26. Liu, Y., Gao, P., Banga, S. & Luo, Z. Q. An *in vivo* gene deletion system for determining temporal requirement of bacterial virulence factors. *Proc. Natl Acad. Sci. USA* **105**, 9385–9390 (2008).
27. Fan, H. Y., Cheng, K. K. & Klein, H. L. Mutations in the RNA polymerase II transcription machinery suppress the hyperrecombination mutant *hpr1Δ* of *Saccharomyces cerevisiae*. *Genetics* **142**, 749–759 (1996).
28. Gietz, R. D., Schiestl, R. H., Willems, A. R. & Woods, R. A. Studies on the transformation of intact yeast cells by the LiAc/SS-DNA/PEG procedure. *Yeast* **11**, 355–360 (1995).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. Isberg for the antibody against SidM and A. Aronson and A. Tao for critical reading of the manuscript and for discussions. This work was supported by NIH-NIAID grants R01AI069344, K02AI085403 and R21AI092043 (Z.-Q.L.).

Author Contributions Y.T. and Z.-Q.L. conceived the project. Y.T. performed the experiments. Y.T. and Z.-Q.L. analysed the data. Z.-Q.L. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to Z.-Q.L. (luoz@purdue.edu).

METHODS

Construction of a *L. pneumophila* genomic library on a yeast expression vector. *L. pneumophila* genomic DNA partially digested with the restriction enzyme *Sau3AI* was separated by agarose gel electrophoresis. DNA fragments of 1–8 kilobase pairs were recovered from the gels and ligated to *Bam*HI-digested yeast vector pGBKT7 (Clontech). Ligated plasmid DNA was introduced into the *E. coli* strain DH5 α by electroporation; approximately 4×10^5 independent colonies were pooled and used to extract total plasmid DNA.

Identification of *L. pneumophila* genes capable of suppressing yeast toxicity of SidM. The open reading frame of *sidM* was inserted into pSB157 (ref. 12) to generate pSB157::SidM, which was digested with *Stu*I and transformed into the yeast strain W303 (ref. 27). The yeast strain W303 (pSB157::SidM), which consistently exhibits galactose-dependent SidM toxicity, was used for the subsequent screenings. Plasmid DNA of the *L. pneumophila* genomic library was transformed into the yeast strain and the transformants were plated onto the selective medium with galactose as the sole carbon source. From about 2×10^6 potential transformants, we obtained a total of 25 colonies that harbour potential suppressor genes. Plasmids carried the potential suppressing genes were rescued and re-introduced into the original yeast strain. Inserts of 16 plasmids that reproducibly suppress the SidM toxicity were sequenced.

Protein purification. To express recombinant proteins, the open reading frames of *sidM*, *sidD*, *rab1* and other *rab* genes were amplified with specific primer pairs (Supplementary Table 1) and were inserted into appropriately digested pGEX6-P-1 or pQE30 (Qiagen), respectively, to produce GST-tagged or His₆-tagged proteins. The cDNAs of other Rab proteins were purchased from Open Biosystems. The integrity of each gene was verified by double-strand sequencing analysis. For protein production, *E. coli* strains harbouring the appropriate expression vector were grown at 37 °C in LB medium (100 μ g ml⁻¹ ampicillin) to a *D*₆₀₀ of 0.5. After adding isopropylthio-D-galactopyranoside (IPTG) to a final concentration of 0.2 mM, the cultures were incubated at 18 °C in a shaker for 18 h. Harvested cells were suspended in PBS buffer and were lysed in a French press at 1,500 p.s.i. The soluble fraction obtained by centrifugation at 6,000g for 10 min at 4 °C was incubated with glutathione Sepharose resin or Ni²⁺ resin (Qiagen) equilibrated with PBS. The proteins were purified as described²⁹. Protein concentrations were determined by the Bradford assay; the purity of all proteins was more than 95% as assessed by SDS-PAGE followed by Coomassie bright blue staining.

Cell culture, infection and transfection. Mouse macrophages were prepared from bone marrow of female A/J mice of 6–10 weeks of age following published protocols¹³. 293T cells were grown in Dulbecco's modified minimum Eagle's medium (DMEM) supplemented with 10% FBS. Established protocols²⁹ were used for transfection and infection.

Cytotoxicity assay. Yeast strains harbouring the P_{gal}::*sidM* constructs co-integrated into the chromosome grown on glucose medium were streaked onto glucose and galactose media, respectively. The growth of the cells was assessed after 3-day incubation at 30 °C. To examine the effects of SidD on the cytotoxicity of SidM1–339 on mammalian cells, we cotransfected 293T cells with plasmids coding for SidD and SidM1–339 at molar ratios of 3 and 5, respectively. Twenty-four hours after transfection, cells expressing the GFP protein were inspected for the cell-rounding phenotype.

In vitro AMPylation and deAMPylation assays. For AMPylation assays using purified recombinant proteins, 1.5 μ g GST-SidM was incubated with 10 μ g of GST-Rab1 for 30 min at 35 °C in an AMPylation buffer containing 25 mM Tris-HCl, pH 7.5, 50 mM NaCl, 3 mM MgCl₂, 0.5 mM EDTA and [α -³²P]ATP (5 μ Ci) (Perkin Elmer). The AMPylation reaction was stopped by the addition of SDS sample buffer. Samples were boiled for 5 min, separated by SDS-PAGE, and detected by autoradiography. The same gels were stained by Coomassie brilliant blue to assess the levels of each protein in the reaction.

To assay the time course of the SidD deAMPylation activity, a 240- μ l master reaction containing 9 μ g His₆-SidM1–339, 60 μ g GST-Rab1 bound on glutathione beads and 30 μ Ci [α -³²P]ATP was first set up and was allowed to proceed for 30 min at 35 °C. We then washed the beads five times in spin columns (Sigma) with 10 \times bed volumes of the AMPylation buffer containing 1% Triton X-100 to remove His₆-SidM1–339, followed by five washes with the AMPylation buffer. This washing regime was effective as no His₆-SidM1–339 was detected by Coomassie brilliant blue staining (Supplementary Fig. 2). Sub-reactions were then established by splitting the master reactions into six different test tubes and 1 μ g His₆-SidD was added to five of these reactions. The reaction that never received His₆-SidD was terminated with SDS sample buffer and used as a control. After incubation at 35 °C for the indicated time duration, the deAMPylation reaction was terminated with SDS loading buffer. To test dose-dependent deAMPylation of Rab1, a master reaction was similarly set up, and after AMPylation, different amounts of SidD (0.6, 1.2, 2.4, 5 μ g) were added into four of the five sub-reactions. The fifth sub-reaction terminated after AMPylation was again used as a control. After incubation at 35 °C for 5 min, the reactions were stopped with SDS loading buffer. Reaction products were separated on 4–20% SDS-PAGE gels (Bio-Rad) and AMPylated GST-Rab1 was detected by autoradiography. Proteins in the gels were detected by Coomassie brilliant blue staining.

Mass spectrometric analysis. A reaction containing 1.5 μ g GST-SidM, 10 μ g GST-Rab1 and 1 mM ATP was allowed to proceed for 30 min at 37 °C. One third of the reaction was withdrawn and terminated with SDS sample buffer to serve as the modified sample. Identical samples were withdrawn at 10 min and 30 min after the addition of 4 μ g of His₆-SidD to the remaining reaction. Proteins were separated by SDS-PAGE and the protein bands corresponding to GST-Rab1 were excised. After complete trypsin digestion, the levels of AMPylation of the samples were analysed by mass spectrometry by the Taplin Mass Spectrometry Facility at Harvard Medical School.

Antibodies, immunostaining and western blot. Antibodies against *Legionella* and GFP were described elsewhere²⁹. The antibodies against Flag (F1804), Rab1 (sc-599) and Myc (sc-40) were purchased from Sigma and Santa Cruz Biotechnology, respectively. The antibody against the yeast metabolic 3-phosphoglycerate kinase (PGK) (a6457) was from Invitrogen. The SidM-specific antibody² was a kind gift from R. Isberg. Cell fixation, permeabilization and immunostaining were performed as described²⁵. The concentrations of antibodies and procedure for western blots were followed established protocols²⁹.

29. Xu, L. *et al.* Inhibition of host vacuolar H⁺-ATPase activity by a *Legionella pneumophila* effector. *PLoS Pathog.* **6**, e1000822 (2010).

Lgr5 homologues associate with Wnt receptors and mediate R-spondin signalling

Wim de Lau^{1*}, Nick Barker^{1†*}, Teck Y. Low², Bon-Kyoung Koo¹, Vivian S. W. Li¹, Hans Teunissen¹, Pekka Kujala³, Andrea Haeggebarth^{1†}, Peter J. Peters³, Marc van de Wetering¹, D. E. Stange¹, J. van Es¹, Daniele Guardavaccaro¹, Richard B. M. Schasfoort⁴, Yasuaki Mohri⁵, Katsuhiko Nishimori⁵, Shabaz Mohammed², Albert J. R. Heck² & Hans Clevers¹

The adult stem cell marker *Lgr5* and its relative *Lgr4* are often co-expressed in Wnt-driven proliferative compartments. We find that conditional deletion of both genes in the mouse gut impairs Wnt target gene expression and results in the rapid demise of intestinal crypts, thus phenocopying Wnt pathway inhibition. Mass spectrometry demonstrates that *Lgr4* and *Lgr5* associate with the Frizzled/Lrp Wnt receptor complex. Each of the four R-spondins, secreted Wnt pathway agonists, can bind to *Lgr4*, -5 and -6. In HEK293 cells, RSP01 enhances canonical WNT signals initiated by WNT3A. Removal of LGR4 does not affect WNT3A signalling, but abrogates the RSP01-mediated signal enhancement, a phenomenon rescued by re-expression of LGR4, -5 or -6. Genetic deletion of *Lgr4/5* in mouse intestinal crypt cultures phenocopies withdrawal of Rspo1 and can be rescued by Wnt pathway activation. *Lgr5* homologues are facultative Wnt receptor components that mediate Wnt signal enhancement by soluble R-spondin proteins. These results will guide future studies towards the application of R-spondins for regenerative purposes of tissues expressing *Lgr5* homologues.

The genes *Lgr4*, *Lgr5* and *Lgr6* encode orphan 7-transmembrane receptors that are close relatives of the receptors for the follicle stimulating hormone (FSH), the luteinizing hormone (LH) and the thyroid-stimulating hormone (TSH)¹. It is currently unknown how they signal and what their ligands are. *Lgr5* is a Wnt target gene which marks proliferative stem cells in several Wnt-dependent stem cell compartments, that is, the small intestine and colon², the stomach³, and the hair follicle⁴. *Lgr6* marks multipotent stem cells in the epidermis⁵. The expression of *Lgr4* is much broader⁶, but we noted that *Lgr5* is co-expressed with *Lgr4* in the stem cell compartments mentioned above. For instance, *Lgr5* marks small intestinal stem cells at the crypt base, whereas *Lgr4* marks all crypt cells, including the *Lgr5*⁺ stem cells (Fig. 1a). The same observation was made in ref. 7. Indeed, our previously reported microarray study, which compared *Lgr5*⁺ stem cells to their immediate daughters, showed that *Lgr4* is expressed in both cell types⁸. Moreover, we generated an allele of *Lgr4* in which a green fluorescent protein-internal ribosomal entry site-CreERT2 (GFP-IRES-CreERT2) cassette was inserted into the ATG start codon, as performed previously for *Lgr5* (ref. 2) and *Lgr6* (ref. 5). Tamoxifen-induced lineage tracing after crossing to the R26R-LacZ Cre reporter strain demonstrated that *Lgr4* is expressed both by short-lived progenitors and by long-lived stem cells (Supplementary Fig. 1). Of note, *Lgr4* and *Lgr5* null mutations are neonatal-lethal in mice^{9,10}.

To address a potential function in crypts, we generated an *Lgr5*^{fl} allele (Supplementary Fig. 2a) by flanking exon 16 with *loxP* sites; its deletion causes a frame shift. This allele was crossed into a mouse strain carrying conditional *Lgr4*^{fl} alleles¹⁰ and the gut-specific *Ah-Cre* transgene, which is inducible by β -naphthoflavone¹¹. Conditional deletion of *Lgr5* alone in the intestinal epithelium of adult mice yielded no apparent phenotype, and did not confirm the Paneth cell phenotype reported previously in *Lgr5* null neonatal mice¹². Deletion of *Lgr4* alone induced loss of proliferation and crypt loss, obvious from day

4–5 post-induction onwards. Paneth cells at crypt bottoms were retained in ‘nests’, disconnected from the surface epithelium. No direct effects were observed on differentiated cells. The combined deletion of *Lgr4* and -5 enhanced this crypt phenotype as judged by the cell proliferation marker Ki67 and the stem cell marker olfactomedin 4 (*Olfm4*; ref. 13). Figure 1b and Supplementary Fig. 2b depict typical results obtained at day 5 post-induction. Of note, hyperplastic wild-type ‘escaper’ crypts (arrows in Fig. 1b, and see ref. 14) served as staining control. Over the next few days, villi shortened and eventually the phenotype was not compatible with life. A gut phenotypic analysis was reported recently for mice homozygous for a hypomorphic *Lgr4* allele that live for about 1 month postnatally⁷. These mice displayed twofold reduced proliferation and an 85% reduction in Paneth cells, whereas stem cell markers and Wnt target genes were unaffected *in vivo*. In ‘minigut’ culture¹⁵, *Lgr4*-hypomorphic crypts failed to initiate organoid growth. No genetic interaction with an *Lgr5* null allele was observed for this intestinal phenotype.

Two signalling pathways, Wnt^{16,17} and Notch¹⁸, are crucial for the maintenance of adult crypt proliferation. We determined gene expression changes by microarraying on day 1 post-deletion of *Lgr4* and -5, when stem cells were physically still present (Supplementary Fig. 3). Simultaneous deletion of *Lgr4* and *Lgr5* resulted in downregulation of 306 genes in two separate experiments (*Lgr4/5* gene set; Supplementary Table 1). These included 29 stem cell-enriched genes⁸. To determine if these are Wnt target genes, we analysed this gene set in two complementary scenarios that detect intestinal Wnt target genes. In the first scenario¹⁹, *in vivo* deletion of *Apc* results in the immediate upregulation of Wnt target genes. Comparison of the *Lgr4/5* gene set to the microarray data from ref. 19 showed a significant upregulation of 53% (137 genes) of the *Lgr4/5* gene set, whereas only 5% (12 genes) were significantly downregulated (Fig. 2a, red ratios). Indeed, gene set enrichment analysis (GSEA)²⁰ showed a highly significant enrichment (false

¹Hubrecht Institute and University Medical Center Utrecht, 3584 CX Utrecht, The Netherlands. ²Biomolecular Mass Spectrometry and Proteomics Group, Bijvoet Center, Utrecht University, 3584 CH Utrecht, The Netherlands. ³Antoni van Leeuwenhoek Hospital Netherlands Cancer Institute, 1066 CX Amsterdam, The Netherlands. ⁴Medical Cell Biophysics, MIRA institute, University of Twente, 7500 AE Enschede, The Netherlands. ⁵Tohoku University, 981-8555 Sendai, Miyagi, Japan. [†]Present addresses: Institute of Medical Biology, 06-06 Immunos, Singapore 138648 (N.B.); Bayer Schering Pharma AG, 13353 Berlin, Germany (A.H.).

*These authors contributed equally to this work.

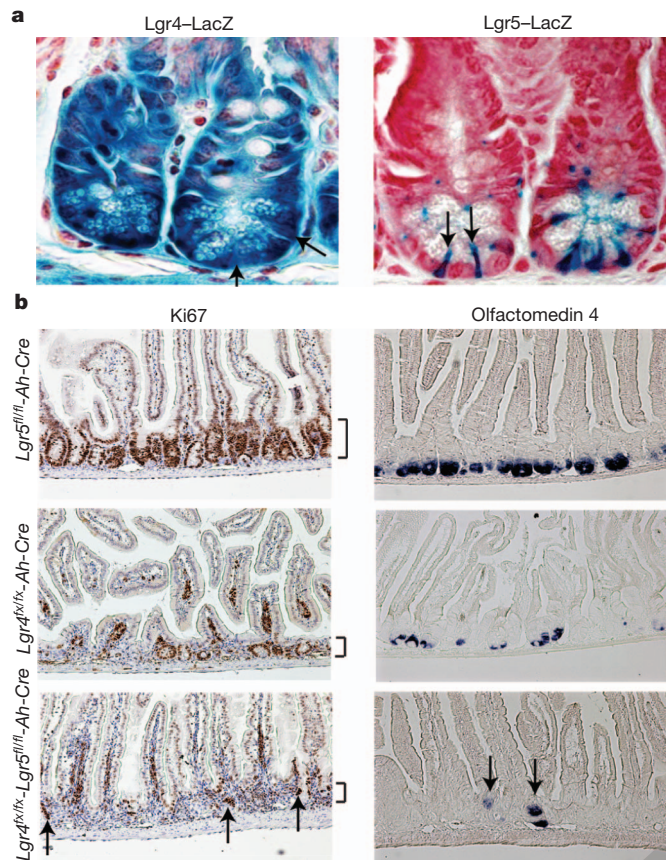


Figure 1 | Conditional deletion of *Lgr4* and *Lgr5*. **a**, Left, expression of *Lgr4*-lacZ² (blue) occurs throughout intestinal crypts. Right, expression of *Lgr5*-lacZ² (blue) is specific to stem cells (arrows) located between Paneth cells at bottom of crypts². **b**, Adult *Lgr4*^{fl/fl} and/or *Lgr5*^{fl/fl} mice carrying the *Ah-Cre* transgene, analysed 5 days after Cre-activation¹¹. Left panels, proliferation visualized by Ki67. Vertical bars, crypt width. Right panels, stem cells visualized by *OlfM4* *in situ* hybridization. Top, *Lgr5* deletion has no obvious effect. Middle, *Lgr4* deletion has significant deleterious effects on crypt stem cells and proliferative progenitors. Bottom, upon *Lgr4/5* double deletion, >80% of crypts entirely disappear. Arrows indicate wild-type escaper crypts that express *OlfM4* and are Ki67⁺.

discovery rate (FDR) and P -value < 0.0001) of the *Lgr4/5* gene set towards the upregulated genes after *Apc* deletion (Fig. 2d). For the second scenario, we performed microarraying before, and 1 day after, acute withdrawal of the Wnt agonist Rspo1 from cultured small intestinal crypt organoids¹⁵. This resulted in the immediate downregulation of 38% (166 genes) of the *Lgr4/5* gene set, whereas only 4% (11 genes) showed the opposite behaviour. (Fig. 2b, green ratios). GSEA analysis confirmed the highly significant enrichment (FDR and P -value < 0.0001) of the *Lgr4/5* gene set towards the downregulated genes after Rspo1 withdrawal (Fig. 2e).

To find molecular partners of LGR receptors, we pursued a tandem affinity purification mass spectrometric strategy. Bait proteins carried double Flag-haemagglutinin (FH)-tags²¹. First, we transiently transfected tagged Frizzled7 (Frz7-FH; frizzled family receptor 7 is also known as FZD7) into HEK293T cells. We detected significant signatures for Frizzled7, and—as expected—the WNT co-receptors LRP5 and LRP6. Surprisingly, multiple peptides were detected for LGR4 (Supplementary Table 2). Of note, we never observed LGR5 homologues, frizzled or LRP proteins in HEK293 cells using ~20 unrelated baits (Supplementary Table 3). We then generated stable clones of LS174T colorectal cancer cells that moderately overexpress tagged versions of LGR4, LGR5 or Frizzled5. The LGR4 bait captured LGR5, LRP6, Frizzled5 and Frizzled7. The LGR5 bait captured

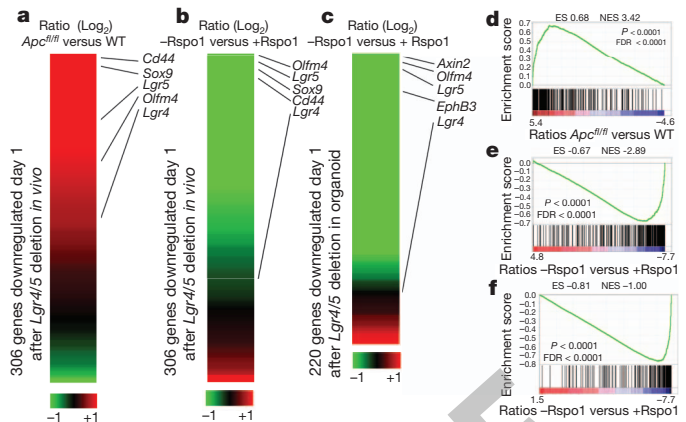


Figure 2 | Wnt target genes are downstream of *Lgr4/5*. Concomitant deletion of *Lgr4* and *Lgr5* in *Lgr4*^{fl/fl} *Lgr5*^{fl/fl} mice resulted in downregulation of 307 unique genes 1 day after deletion (Supplementary Table 1). The intestinal Wnt signature can be revealed by two opposing experiments: deletion of *Apc* *in vivo* results in upregulation of a Wnt target signature¹⁹, whereas Rspo1 withdrawal from intestinal organoids¹⁵ results in downregulation of this signature. **a**, Heat map of the log₂ ratio of *Apc*^{fl/fl} mice versus control wild-type mice (WT) for the 307 *Lgr4/5* genes 3 days after deletion of *Apc* (ratios taken from ref. 19). **b**, Heat map of the log₂ ratio of intestinal organoids 1 day after Rspo1 withdrawal (–Rspo1) versus control organoids (+Rspo1) for the 307 *Lgr4/5* gene set. **c**, Heat map of the log₂ ratio of intestinal organoids 1 day after Rspo1 withdrawal (–Rspo1) versus control organoids (+Rspo1) for the 220 downregulated genes in *Lgr4*^{fl/fl} *Lgr5*^{fl/fl} organoids (1 day after deletion). **d**, Gene set enrichment analysis (GSEA). Genes are ranked according to their differential expression between *Apc*-deleted and wild-type mice (data from ref. 19). Black bars beneath the graph depict the rank positions of the 306 genes from the *Lgr4/5* gene set. A highly significant enrichment of the 306 *Lgr4/5* genes was detected towards the gene set upregulated 3 days after *Apc* deletion *in vivo*. **e**, Genes ranked according to differential expression in intestinal organoids after Rspo1 withdrawal versus control organoids. GSEA shows a highly significant enrichment of the 306 *Lgr4/5* genes towards the genes downregulated after Rspo1 withdrawal. **f**, Genes are ranked according to their differential expression in intestinal organoids after Rspo1 withdrawal versus control organoids. GSEA shows a highly significant enrichment of the 220 *Lgr4/5* downregulated genes in organoids towards those genes downregulated after Rspo1 withdrawal. ES, enrichment score; FDR, false discovery rate; NES, normalized ES.

Frizzled6 and LRP5 and LRP6. The Frizzled5 bait captured LGR4, LGR5, LRP5 and LRP6. These proteins were never observed in the non-transfected controls run in parallel. Results are summarized in Supplementary Table 2. Thus, LGR4 and LGR5 can occur in a physical complex with frizzled proteins and LRP5/6.

The four secreted R-spondin proteins activate the canonical Wnt pathway and are particularly potent when synergizing with secreted Wnt proteins^{22,23}. Systemic delivery of Rspo1 in mice leads to a dramatic enhancement of the Wnt-responsive intestinal crypts²⁴ and stimulates epithelial repair²⁵. Whereas Rspo3 uses syndecans as receptors to mediate non-canonical Wnt signals²⁶, the R-spondin receptors that drive canonical Wnt signals have remained controversial²⁶. Rspo1 was proposed to bind the Wnt co-receptor Lrp6 (refs 27, 28), to block the Kremen protein that downregulates surface expression of Wnt receptors²⁹, or to block the interaction of the Wnt inhibitor Dkk1 with Lrp6 (ref. 23). HEK293T cells are responsive to WNT3A and R-spondin²². We used this cell line to search for the cognate R-spondin receptor. To validate the efficacy of using secreted baits to detect surface receptors, we generated a tagged version of Dkk1 (Dkk1-FH). Dkk1 interacts with LRP5/6 (ref. 30). We incubated 2×10^9 HEK293T cells with Dkk1-FH. The cells were washed, lysed and immunoprecipitated with an anti-Flag antibody. The immunoprecipitated material was eluted with Flag peptide, and re-precipitated with anti-haemagglutinin antibody, after which mass spectrometric analysis was performed. We readily identified Dkk1 and endogenous

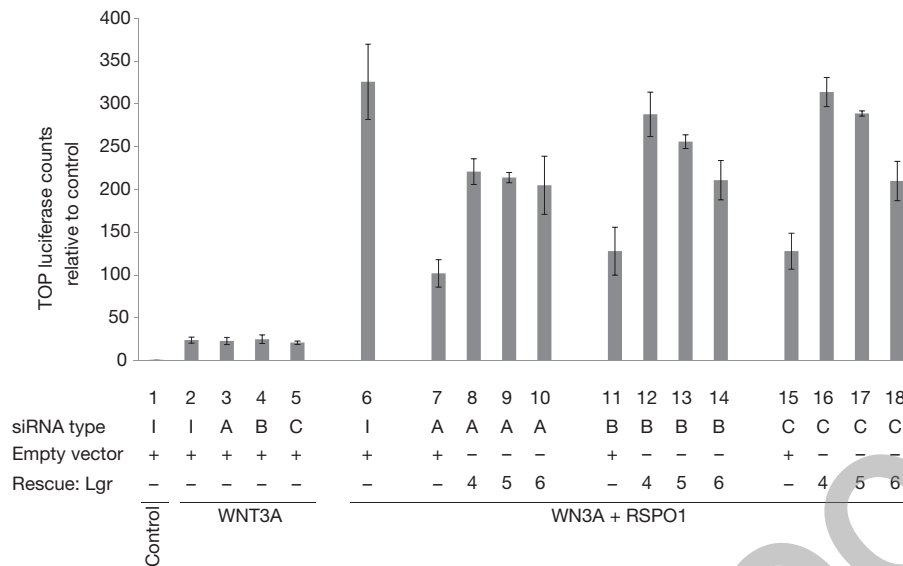


Figure 3 | LGR4 is essential for transmitting RSPO1 signals but dispensable for transmitting WNT3A signals. TOPFLASH Wnt reporter assays. HEK293T cells were transfected with indicated siRNAs. siRNAs A, B, C, siRNAs directed at the 3'UTR of human *LGR4*; siRNA I, pool of scrambled siRNA. Three days later, the cells were transfected with TOPFLASH reporters and 5 ng of the indicated rescue constructs, and incubated with WNT3A and

RSPO1 as indicated. WNT3A induced TOPFLASH reporter activity without showing effects of removal of LGR4. R-spondin potentiated the WNT3A response, but this effect was sensitive to removal of LGR4. Rescue was obtained with LGR4, -5, and -6 rescue constructs, but not with control vector (lanes 7, 11, 15). $n = 4$. Error bars indicated as s.d.

LRP5 and LRP6 (Supplementary Table 2). We then produced a double-tagged version of human RSPO1 (RSPO1-FH) at $\sim 1 \mu\text{g ml}^{-1}$. As tested in the TOPFLASH WNT reporter assay³¹, the protein potentiated the effect of Wnt3A by approximately 12-fold (Fig. 3). We incubated 2×10^9 HEK293 cells with the RSPO1-FH-conditioned medium and subjected them to the above protocol. The only transmembrane protein detected was LGR4 (Supplementary Table 2).

HEK293T cells were transiently transfected with tagged versions of LGR4, LGR5 and LGR6, and with the closely related LGR1 (FSH

receptor), LGR3 (TSH receptor), LGR7 and LGR8. Transfected cells were incubated with Fc fusion protein of human RSPO1 at $\sim 1 \mu\text{g ml}^{-1}$, washed, lysed and precipitated with protein G beads. Western blotting for the Flag tag revealed binding of RSPO1-Fc to LGR4, LGR5 and LGR6, but not to the other LGR proteins (Fig. 4a). In the same assay, Fc fusion proteins of RSPO2, RSPO3 and RSPO4 at $\sim 1 \mu\text{g ml}^{-1}$ were shown to bind specifically to LGR4, LGR5 and LGR6 (Supplementary Fig. 4).

Soluble RSPO1-FH interacted with the leucine-rich-repeat exodomain of LGR5 (amino acids 1 to 546) expressed as a human IgG-Fc fusion protein (LGR5-exo-Fc). This allowed surface plasmon resonance array imaging. Anti-Flag antibody, spotted on the sensor chip, captured RSPO1-FH (Fig. 4b, left, blue line). LGR5-exo-Fc bound to RSPO1-FH (left; blue line). After regeneration, a control experiment was performed using a noggin-Fc fusion protein (Fig. 4b, right, blue line), which did not bind RSPO1-FH. Both LGR5-exo-Fc and noggin-Fc could be captured on anti-human IgG spotted as a control (Fig. 4b, red line). The K_D of the LGR5-exo-Fc-RSPO1-FH interaction was determined at 3.1 nM by an extrapolated 1:1 interaction model at serial injections of RSPO1 and LGR5-exo-Fc on anti-Flag spots with decreasing densities (Supplementary Fig. 5). In a dose-response curve in the HEK293T Wnt reporter assay, Wnt signalling was indeed induced by nanomolar amounts of RSPO1 (Supplementary Fig. 6).

Rat monoclonal antibodies were raised against full-length human LGR5 protein. All antibodies reacted with LGR5-exo-Fc and their

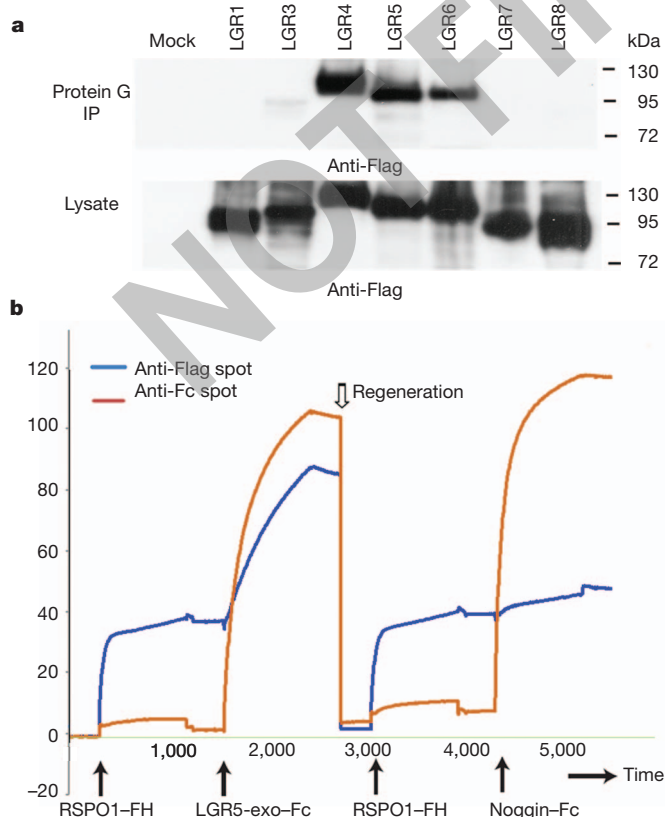


Figure 4 | Direct physical interaction of RSPO1 with LGR4/5/6 exodomains. **a**, RSPO1 binds to LGR4, LGR5 and LGR6. Top panel, HEK293T cells transfected with Flag-tagged versions of the indicated LGR proteins, incubated with RSPO1-Fc fusion protein at $\sim 1 \mu\text{g ml}^{-1}$. Cells were washed, lysed and RSPO1-Fc was immunoprecipitated with protein G beads. Top, western blotting for Flag revealed specific binding to LGR4, LGR5 and LGR6. IP, immunoprecipitation. Bottom, input of tagged proteins. **b**, RSPO1-LGR5 interaction visualized by surface plasmon resonance. Anti-Flag antibody spotted on the sensor chip captured RSPO1-FH (Fig. 4, left, blue line). After a wash, LGR5-exo-Fc bound to RSPO1-FH. After regeneration with low pH, a noggin-Fc fusion protein served as negative control (Fig. 3, right, blue line). Both LGR5-exo-Fc and noggin-Fc could be captured on goat anti-human IgG spotted as a control (Fig. 3, red line).

epitopes were mapped using carboxy-terminal deletion clones and human-mouse hybrids of the *Lgr5* extracellular domain (Supplementary Fig. 7a). All antibodies recognizing the amino terminus of the *Lgr5* extracellular domain blocked the RSPO1 interaction, whereas the other antibodies did not (Supplementary Fig. 7b). Indeed, when the extreme N terminus or the first LRR domain were deleted from *Lgr5*, RSPO1-binding was lost (Supplementary Fig. 8).

HEK293T cells express *LGR4* but not *LGR5* (Supplementary Fig. 9). To test if *LGR4* constitutes a functional RSPO1 receptor, we removed *LGR4* mRNA from HEK293T cells with three independent small interfering RNAs (siRNAs; Supplementary Fig. 9). Subsequently, the cells were transfected with TOPFLASH WNT reporters and their

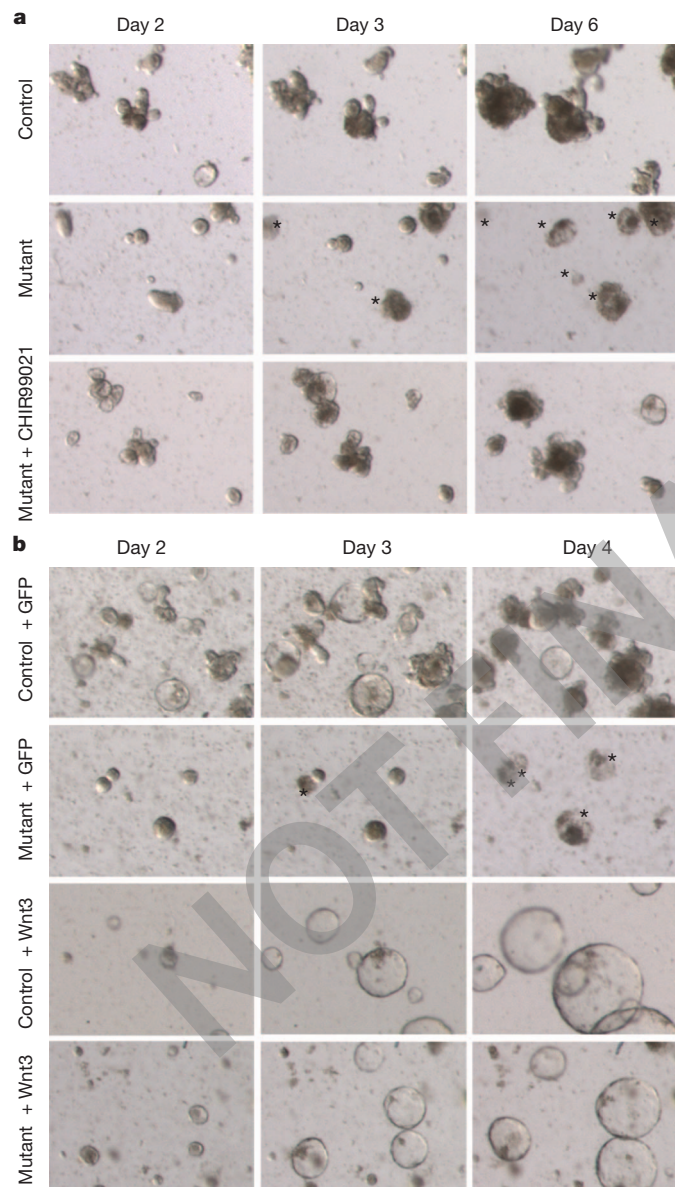


Figure 5 | Rescue of *Lgr4/5* deletion in cultured crypt organoids by Wnt signals. **a**, Organoids established under standard conditions¹⁵ from *Lgr4*^{fl/fl} *Lgr5*^{fl/fl} Villin-CreERT2 mice³⁷ (Mutant) or from control Villin-CreERT2 mice (Control). Tamoxifen treatment leads to death (asterisks) of mutant organoids but not that of controls. This is overcome by addition of CHIR99021 at 5 μ M (Mutant + CHIR99021). **b**, Organoids established from *Lgr4*^{fl/fl} *Lgr5*^{fl/fl} Villin-CreERT2 mice (Mutant) or from control Villin-CreERT2 mice (Control) are infected with Wnt3-expressing retrovirus which turns organoids into growing, rounded cysts³⁴, or with control (GFP) retrovirus. Subsequent deletion of *Lgr4* and -5 has no effect on Wnt3-expressing organoids but leads to death of control retrovirus organoids.

response to exogenously added WNT3A and RSPO1 was measured. WNT3A alone induced a ~25 fold increase in TOPFLASH activity. Removal of *LGR4* had no effect on this WNT3A response (Fig. 3; compare bar 1 to bars 2–5). The combination of WNT3A with RSPO1 led to a further ~12-fold increase in TOPFLASH activity (bar 6). This effect was greatly diminished when *LGR4* was removed (bars 7, 11 and 15), but could be rescued by transfection of *LGR4* (bars 8, 12 and 16), *LGR5* (bars 9, 13 and 17) or *LGR6* expression plasmids (bars 10, 14 and 18). No rescue was obtained with *LGR1*, *LGR7* or *LGR8* (Supplementary Fig. 10). A ‘DRY’ motif at the cytosolic end of the third transmembrane region is believed to be essential for protein G coupling in serpentine receptors³². We mutated the ERG motif at the corresponding position in *LGR5* to ENG. Both versions of *LGR5* rescued loss of *LGR4* in the TOPFLASH assay (Supplementary Fig. 11), indicating that protein G signalling is not involved in signalling towards β -catenin.

The proposed model implies that strong Wnt signals should rescue loss of *Lgr* receptors. We generated intestinal organoids¹⁵ from *Lgr4*^{fl/fl} *Lgr5*^{fl/fl} \times Villin-CreERT2 mice. Tamoxifen-induced deletion *in vitro* resulted in the demise of the organoids after 2 days (Fig. 5a). Microarraying was performed 1 day after induction of deletion, that is, 24h before organoids lost viability. This revealed a striking similarity between gene expression changes 1 day after withdrawal of Rspo1 and 1 day after deletion of the proposed R-spondin receptors *Lgr4* and *Lgr5* (Fig. 2c and 2f). Addition of the Gsk3 inhibitor CHIR99021 (which activates the Wnt cascade) led to a quantitative, persistent rescue of *Lgr4/5*-mutant organoids (Supplementary Fig. 12). Of note, LiCl (which inhibits Gsk3, but also inositol monophosphatase³⁴) reportedly allows 6% rescue of the growth of *Lgr4*-hypomorphic organoids⁷. We next expressed Wnt3 (ref. 35) by retroviral transduction in established *Lgr4*^{fl/fl} *Lgr5*^{fl/fl} \times Villin-CreERT2. Subsequent tamoxifen-induced deletion of *Lgr4* and -5 lead to the demise of the mutant crypts, but this effect was robustly rescued by Wnt3 overexpression (Fig. 5b). In a similar approach, retrovirally expressed *Lgr5* and *Lgr5*-ENG rescued organoid growth (Supplementary Fig. 13).

These data demonstrate that binding of R-spondins to their cognate receptors, the *Lgr5* homologues, triggers downstream canonical Wnt signals through associated frizzled-Lrp5/6 complexes. The current observations assign a crucial function to *Lgr4* and *Lgr5* in Wnt-dependent stem cells and progenitor cells: In intestinal crypts, *Lgr4* and -5, incorporated into frizzled-Lrp complexes, allow R-spondins to augment short-range Wnt signals emanating from Paneth cells³⁴. This provides a molecular mechanism for the potent hyperplastic response of crypts to Rspo1 (ref. 24) and will guide future studies towards the application of R-spondins for regenerative purposes of tissues expressing *Lgr5* homologues. Moreover, this study reinforces the connection, first described in intestinal crypts³⁶, between Wnt signalling and adult mammalian stem cell biology.

METHODS SUMMARY

Mice. Experiments performed according to guidelines and reviewed by the DEC of the KNAW. Knock-in alleles were generated and conditional deletion was induced as described elsewhere^{2,13}.

Plasmid expression constructs. Available upon request.

Histology. Tissues were prepared and stained with Ki67, PAS, or subjected to fluorescent *in situ* hybridization to *Olfm4* as described elsewhere¹³.

Cell culture, transfections, TOPFLASH assays, RSPO1/WNT3A stimulation. HEK293T cells and Ls174T cells were cultured and transfected as described elsewhere^{31,36}. siRNAs directed at the 3' untranslated repeat (3'UTR) of human *LGR4* were from Thermo Scientific Dharmacon. Sense sequences: A, 5'-gaaagaa aacuguggaauu-3'; B, 5'-gggaggagucuaaaguuuu-3'; C, 5'-gguaagaacuccu aauuuuu-3'; I (irrelevant), ON-TARGETplus Non-Targeting Pool from Dharmacon. Transfected with Dharmafect1.

Microarraying. Performed on an Agilent platform, as described elsewhere³.

Immunoprecipitation, epitope mapping, western blotting. Performed using standard protocols. Mouse M2 Anti-Flag antibody (Sigma), rabbit anti-human IgG-Fc horse radish peroxidase (HRP) conjugate (Pierce), and goat anti-rat IgG-HRP-conjugate (Pierce) were used. Protein G agarose beads (Millipore).

Mass spectrometric analysis. In all cases, samples were obtained by sequential immunoprecipitation, performed using anti-Flag affinity matrix (Sigma), 3×Flag peptide (Sigma) elution and re-precipitation using anti-haemagglutinin affinity matrix (Roche) as described²¹. For a detailed description of the mass spectrometric analysis see Supplementary Material.

Surface plasmon resonance. Performed by IBIS Technologies, Enschede, The Netherlands.

Rat monoclonal antibodies anti-human LGR5. Generated by Genovac by DNA vaccination with full-length human *LGR5* cDNA. Initial hybridoma screening was performed by fluorescence-activated cell sorting (FACS) on stable human *LGR5* transfectants.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 29 March; accepted 27 June 2011.

Published online 4 July 2011.

- Barker, N. & Clevers, H. Leucine-rich repeat-containing G-protein-coupled receptors as markers of adult stem cells. *Gastroenterology* **138**, 1681–1696 (2010).
- Barker, N. *et al.* Identification of stem cells in small intestine and colon by marker gene *Lgr5*. *Nature* **449**, 1003–1007 (2007).
- Barker, N. *et al.* *Lgr5*⁺ stem cells drive self-renewal in the stomach and build long-lived gastric units *in vitro*. *Cell Stem Cell* **6**, 25–36 (2010).
- Jaks, V. *et al.* *Lgr5* marks cycling, yet long-lived, hair follicle stem cells. *Nature Genet.* **40**, 1291–1299 (2008).
- Snippert, H. J. *et al.* *Lgr6* marks stem cells in the hair follicle that generate all cell lineages of the skin. *Science* **327**, 1385–1389 (2010).
- Van Schoore, G., Mendive, F., Pochet, R. & Vassart, G. Expression pattern of the orphan receptor *LGR4/GPR48* gene in the mouse. *Histochem. Cell Biol.* **124**, 35–50 (2005).
- Mustata, R. C. *et al.* *Lgr4* is required for Paneth cell differentiation and maintenance of intestinal stem cells I. *EMBO Reports* **12**, 558–564 (2011).
- van der Flier, L. G. *et al.* Transcription factor achaete scute-like 2 controls intestinal stem cell fate. *Cell* **136**, 903–912 (2009).
- Morita, H. *et al.* Neonatal lethality of *LGR5* null mice is associated with ankyloglossia and gastrointestinal distension. *Mol. Cell. Biol.* **24**, 9736–9743 (2004).
- Kato, S. *et al.* Eye-open at birth phenotype with reduced keratinocyte motility in *LGR4* null mice. *FEBS Lett.* **581**, 4685–4690 (2007).
- Ireland, H., Houghton, C., Howard, L. & Winton, D. J. Cellular inheritance of a Cre-activated reporter gene to determine Paneth cell longevity in the murine small intestine. *Dev. Dyn.* **233**, 1332–1336 (2005).
- Garcia, M. I. *et al.* *LGR5* deficiency deregulates Wnt signaling and leads to precocious Paneth cell differentiation in the fetal intestine. *Dev. Biol.* **331**, 58–67 (2009).
- van der Flier, L. G., Haeghebarth, A., Stange, D. E., van de Wetering, M. & Clevers, H. OLFM4 is a robust marker for stem cells in human intestine and marks a subset of colorectal cancer cells. *Gastroenterology* **137**, 15–17 (2009).
- Muncan, V. *et al.* Rapid loss of intestinal crypts upon conditional deletion of the Wnt/Tcf-4 target gene *c-Myc*. *Mol. Cell. Biol.* **26**, 8418–8426 (2006).
- Sato, T. *et al.* Single *Lgr5* stem cells build crypt-villus structures *in vitro* without a mesenchymal niche. *Nature* **459**, 262–265 (2009).
- Pinto, D., Gregorieff, A., Begthel, H. & Clevers, H. Canonical Wnt signals are essential for homeostasis of the intestinal epithelium. *Genes Dev.* **17**, 1709–1713 (2003).
- Kuhnert, F. *et al.* Essential requirement for Wnt signaling in proliferation of adult small intestine and colon revealed by adenoviral expression of Dickkopf-1. *Proc. Natl Acad. Sci. USA* **101**, 266–271 (2004).
- van Es, J. H. *et al.* Notch/γ-secretase inhibition turns proliferative cells in intestinal crypts and adenomas into goblet cells. *Nature* **435**, 959–963 (2005).
- Sansom, O. J. *et al.* *Myc* deletion rescues *Apc* deficiency in the small intestine. *Nature* **446**, 676–679 (2007).
- Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
- Nakatani, Y. & Ogryzko, V. Immunoaffinity purification of mammalian protein complexes. *Methods Enzymol.* **370**, 430–444 (2003).
- Kazanskaya, O. *et al.* R-Spondin2 is a secreted activator of Wnt/β-catenin signaling and is required for *Xenopus* myogenesis. *Dev. Cell* **7**, 525–534 (2004).
- Kim, K. A. *et al.* R-spondin family members regulate the Wnt pathway by a common mechanism. *Mol. Biol. Cell* **19**, 2588–2596 (2008).
- Kim, K. A. *et al.* Mitogenic influence of human R-spondin1 on the intestinal epithelium. *Science* **309**, 1256–1259 (2005).
- Zhao, J. *et al.* R-spondin1, a novel intestinotrophic mitogen, ameliorates experimental colitis in mice. *Gastroenterology* **132**, 1331–1343 (2007).
- Ohkawara, B., Glinka, A. & Niehrs, C. Rspo3 binds syndecan 4 and induces Wnt/PCP signaling via clathrin-mediated endocytosis to promote morphogenesis. *Dev. Cell* **20**, 303–314 (2011).
- Nam, J. S., Turcotte, T. J., Smith, P. F., Choi, S. & Yoon, J. K. Mouse cristin/R-spondin family proteins are novel ligands for the Frizzled 8 and LRP6 receptors and activate β-catenin-dependent gene expression. *J. Biol. Chem.* **281**, 13247–13257 (2006).
- Wei, Q. *et al.* R-spondin1 is a high affinity ligand for LRP6 and induces LRP6 phosphorylation and β-catenin signaling. *J. Biol. Chem.* **282**, 15903–15911 (2007).
- Binnerts, M. E. *et al.* R-Spondin1 regulates Wnt signaling by inhibiting internalization of LRP6. *Proc. Natl Acad. Sci. USA* **104**, 14700–14705 (2007).
- Glinka, A. *et al.* Dickkopf-1 is a member of a new family of secreted proteins and functions in head induction. *Nature* **391**, 357–362 (1998).
- Korinek, V. *et al.* Constitutive transcriptional activation by a β-catenin-Tcf complex in APC^{−/−} colon carcinoma. *Science* **275**, 1784–1787 (1997).
- Flanagan, C. A. A GPCR that is not “DRY”. *Mol. Pharmacol.* **68**, 1–3 (2005).
- Robine, S., Sahuguillo-Merino, C., Louvard, D. & Pringault, E. Regulatory sequences on the human villin gene trigger the expression of a reporter gene in a differentiating HT29 intestinal cell line. *J. Biol. Chem.* **268**, 11426–11434 (1993).
- Klein, P. S. & Melton, D. A. A molecular mechanism for the effect of lithium on development. *Proc. Natl Acad. Sci. USA* **93**, 8455–8459 (1996).
- Sato, T. *et al.* Paneth cells constitute the niche for *Lgr5* stem cells in intestinal crypts. *Nature* **469**, 415–418 (2011).
- Korinek, V. *et al.* Depletion of epithelial stem-cell compartments in the small intestine of mice lacking Tcf-4. *Nature Genet.* **19**, 379–383 (1998).
- Ng, S. S. *et al.* Phosphatidylinositol 3-kinase signaling does not activate the Wnt cascade. *J. Biol. Chem.* **284**, 35308–35313 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank G. Vassart for *Lgr4-LacZ* intestinal tissue, D. Winton for *Ah-Cre* mice, S. Robine for *Villin-CreERT2* mice, A. Moerkamp and C. Verheul for experimental help and H. Farin for figures.

Author Contributions All Hubrecht Institute authors performed experiments under guidance of H.C.; S.M., A.J.R.H. and T.Y.L. performed mass spectrometry; P.K. and P.J.P. performed electron microscopy analysis; R.B.M.S. performed plasmon surface resonance; and Y.M. and K.N. generated the *Lgr4* knockout mouse.

Author Information Microarray data have been deposited in the GEO database under accession number GSE28265. Mass spectrometry data sets are available at ProteomeCommons.org Tranche Repository <https://proteomecommons.org/tranche/data-downloader.jsp?h=2LOW5tCJBOfT%2FpcCatMrPqCgTT0d247s6poPgSvwu16KiVwCfExWdJOfGd14FraidTHUn1PYHoToNts1zdwKmKEAAAAAAACzw%3D%3D>. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to H.C. (h.clevers@hubrecht.eu).

METHODS

Microarray analysis *Lgr4/5* knockout mice. Small intestinal crypts were isolated 1 day after induction of deletion from *AHCre Lgr4^{flx/flx} Lgr5^{fl/fl}* as well as wild-type mice by incubation in 2 mM EDTA. RNA was isolated using TRIzol (Invitrogen) and 1 µg of RNA was labelled using a Quick Amp Labelling Kit, two colour (Agilent Technologies) with Cy5 and Cy3, respectively; all as described elsewhere³. Two separate biological replicates were performed in dye-swap, resulting in four individual arrays. Labelling, hybridization and washing were done according to Agilent guidelines. Differentially labelled cRNA was hybridized on 4X44K Agilent Whole Mouse Genome dual colour Microarrays (G4122F). Array data were normalized and retrieved using Feature Extraction (V.9.5.3, Agilent Technologies) and data analyses were performed using Microsoft Excel (Microsoft Corporation). Features were flagged, if signal intensities for both the Cy3 and Cy5 channel did not pass the feature extraction filter, 'significant and positive' or 'well above background'. Genes were considered downregulated if 4 out of 4 arrays showed a significant ($P < 0.05$) downregulation of < -0.58 (linear -1.5 fold). This resulted in 379 entries, which 307 unique genes (Supplementary Table array data will be available at the Gene Expression Omnibus <http://www.ncbi.nlm.nih.gov/geo>).

R-spondin1 withdrawal in mouse intestinal organoids. Crypts were isolated from a wild-type mouse small intestine by incubating with 2 mM EDTA in PBS for 30 min at 4 °C, and subsequently grown in crypt culture medium as reported previously¹⁶. Briefly, isolated crypts were cultured in Matrigel (BD Bioscience) in 24-well plates, and advanced DMEM/F12 medium (Invitrogen) containing epithelial growth factor, noggin and Rspo1 was added after polymerization of Matrigel. Confluent organoids were split into multiple wells, and were then cultured in crypt culture medium in the presence or absence of Rspo1. One day after Rspo1 withdrawal organoids (−Rspo1 organoids) and the control organoids (+Rspo1 organoids) were then collected for RNA extraction and microarray analysis. RNA from control and Rspo1 depleted organoids (1 µg), together with universal mouse reference RNA (Stratagene), was labelled using a Quick Amp Labelling Kit, two colour (Agilent Technologies) with Cy5 and Cy3, respectively. Samples were hybridized to 4X44K Whole Mouse Genome Microarrays (Agilent, G4122F) according to manufacturer's instructions. Microarray signal and background information were retrieved and normalized using the Feature Extraction program (V.9.5.3, Agilent Technologies). Samples were considered as well-measured when the fluorescent signals in red channel (Cy5) in either of the samples were greater than twofold above the local background. Differences between −Rspo1 organoids and +Rspo1 organoids were calculated by subtracting the ratio for '−Rspo1 organoids versus reference RNA' from '+Rspo1 versus reference RNA'. Array data is available at the Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo>).

Heat map and gene set enrichment analysis. Heat maps were generated using MeV (Multiple Experiment Viewer v.4.3)⁴². Heat maps were generated using the 306 genes from the *Lgr4/5* gene set and plotting the ratios of two different experiments for these genes. Figure 2a contains the ratios from ref. 15, where the authors deleted *Apc* in the *Ah-Cre Apc^{fl/fl}* mice and performed microarray analysis 3 days after deletion. Figure 2b contains the ratios from the Rspo1 withdrawal experiment described above. Gene set enrichment analysis (GSEA) implemented with GSEAP v.2.0 (<http://www.broad.mit.edu/gsea>) was used to identify significant enrichments of the *Lgr4/5* gene set in the two different experimental scenarios mentioned above. All 'well measured' features ($n = 20,844$) for the Rspo1 experiment from the Agilent arrays and all features with a ratio ($n = 45,101$) for the *Apc* arrays from the Affymetrix platform were used for the ranked gene list in GSEA.

Mass spectrometric analysis. The samples were subjected to SDS-PAGE followed by Coomassie blue staining. The gel lane was sliced into 24 equal sections and digested. In brief, protein reduction and alkylation was performed with dithiothreitol (60 °C, 1 h) and Iodoacetamide (dark, room temperature, 30 min), respectively. Digestion was performed with trypsin overnight at 37 °C. Peptides were extracted with 10% formic acid³⁸.

The extracted peptides were subjected to nanoscale liquid chromatography tandem mass spectrometry (nanoLC-MS/MS) analysis, performed on an Agilent 1200 HPLC system (Agilent technologies) connected to an LTQ Orbitrap Velos (ThermoFisher)³⁹. The nanoLC was equipped with a 20 mm × 100 µm internal diameter trap column and a 400 mm × 50 µm internal diameter analytical column (Reprosil C18, Dr Maisch). Trapping was performed at a flow of 5 µl min^{−1} for 10 min and the fractions were eluted using a 45 min linear gradient from 0 to 40% solvent B (0.1 M acetic acid in 80% acetonitrile (v/v), in which solvent A was 0.1 M acetic acid), 40 to 100% solvent B in 2 min and 100% B for 2.5 min. The analytical flow rate was 100 nl min^{−1} and the column effluent was directly introduced into the electrospray source of the mass spectrometer using a standard coated fused silica emitter (New Objective) (outer diameter 360 µm, tip internal diameter

10 µm) biased to 1.7 kV. The mass spectrometer was operated in positive ion mode and in data-dependent mode to automatically switch between MS and MS/MS. The five most intense ions in the survey scan were fragmented in the linear ion trap using collision-induced dissociation⁴⁰. The target ion setting was 5×10^5 for the Orbitrap, with a maximum fill-time of 250 ms. Fragment ion spectra were acquired in the LTQ with an automatic gain control value of 5×10^3 and a max injection time of 100 ms.

Protein identification: raw mass spectrometry data were converted to peak lists using MaxQuant version 1.0.13.13 (ref. 41). Spectra were searched against the IPI (International Protein Index) Human database version 3.37 (69,164 sequences; 29,064,824 residues) using the Mascot search engine (version 2.3.02; www.matrixscience.com), with trypsin set as enzyme. The database search was made with the following parameters set to consider a peptide tolerance of ± 15 p.p.m., a fragment tolerance of ± 0.5 Da, allowing two missed cleavages, carbamidomethyl (C) as fixed modification; oxidation (M) and protein N-terminal acetylation as variable modifications.

Plasmon surface resonance. A pre-activated ester sensor chip (IBIS Technologies) was spotted using a Continuous Flow Microspotter (Wasatch Microfluidics). In total, 32 spots were created with both mouse anti-Flag and goat anti-human IgG in 8 × serial dilution (start concentrations were 500 µg ml^{−1} and 100 µg ml^{−1}, respectively) in 10 mM MES buffer pH 5.5. Control (reference) spots contained human serum albumin, anti-human serum albumin and MES buffer. After preparing the sensor chip, the slider was positioned in an IBIS MX96 (IBIS Technologies) for label-free surface plasmon resonance (SPR) array analysis. The instrument enables multiplex interactions up to 96-plex using scanning imaging optics for automated calculation of the SPR dip shifts of all region of interests simultaneously. The signal-to-noise ratio of the instrument which reflects the limit of detection is better than 1 RU corresponding to 1 picogram protein per square millimetre. In the IBIS MX96, back and forth mixing is applied enabling minimal use of sample while the length of the exposure of the sample to the microarray is unlimited and not affected by the volume of, for example, an injection loop. A two-step interaction process was carried out and the multiplex interaction event to all spots of the array was recorded simultaneously. In this way, chip-to-chip and sample-to-sample variations can be excluded, while positive and negative controls and referencing can be applied instantly. A microscope image of the sensor chip (not shown here) can reveal any irregularities, heterogeneities of the spots and/or disturbing air-bubbles. First 70 µl of RSP01-FH was injected followed by dissociation and second injection of 70 µl LGR5-Fc. After regeneration with acid buffer (Gly-HCl, 10 mM, pH 2.0) for 120 s, the control experiment was capture of RSP01-FH followed by injection of noggin-Fc (results see Fig. 4b). For affinity constant determination first RSP01-FH was captured on the anti-Flag spot until saturation of the spot followed by injection of LGR5-Fc (8 µg ml^{−1}). RSP01 was not immobilized directly on the chip because it did not survive the acid regeneration process (data not shown). Capturing of RSP01-FH was possible on the anti-Flag antibody spot from growth medium. Supplementary Fig. 1 is an overlay plot of three interaction series of first RSP01-FH (2 min association was sufficient to reach saturation) and three different concentrations of LGR5-Fc (molecular mass 176 kDa) corresponding to 45 nM, 23 nM and 11 nM injections. Referencing was carried out by subtraction of the anti-Flag spot signal with the signal coming from a HAS-loaded spot in SPRint software (IBIS Technologies) for compensating bulk refractive index shifts for example, as a result of growth medium compounds. The affinity constant was calculated using a discrete 1:1 interaction model using global fitting (Scrubber 2, BioLogic Software Pty Ltd). In Supplementary Fig. 1, the residual plot of the experimental curve minus fit values is shown and although it cannot be revealed that the interaction process is according to a discrete 1:1 interaction model the affinity constant was calculated to be ~3 nM. It was not necessary to add an extra fit parameter for mass transport limitation compensation, due to a high back and forth mixing condition in the IBIS MX96 flow cell.

38. Shevchenko, A., Wilm, M., Vorm, O. & Mann, M. Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal. Chem.* **68**, 850–858 (1996).
39. Rajmakers, R. *et al.* Automated online sequential isotope labeling for protein quantitation applied to proteasome tissue-specific diversity. *Mol. Cell. Proteomics* **7**, 1755–1762 (2008).
40. Frese, C. K. *et al.* Improved peptide identification by targeted fragmentation using CID, HCD and ETD on a LTQ-Orbitrap Velos. *J. Proteome Res.* **10**, 2377–2388 (2011).
41. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnol.* **26**, 1367–1372 (2008).
42. Saeed, A. I. *et al.* TM4: a free, open-source system for microarray data management and analysis. *Biotechniques*, **34**, 374–378 (2003).